AN

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 August 2002 (01.08.2002)

PCT

(10) International Publication Number
WO 02/059151 A2

(51) International Patent Classification[7]: C07K 14/705, 16/28, C12N 15/12, 5/10, C12Q 1/68, A01K 67/027, A61K 38/17, G01N 33/566, 33/68

(21) International Application Number: PCT/US01/42978

(22) International Filing Date:
6 November 2001 (06.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/769,741 26 January 2001 (26.01.2001) US

(71) Applicant: PE CORPORATION (NY) [US/US]; 761 Main Avenue, Norwalk, CT 06859 (US).

(72) Inventors: LI, Zhenya; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). CRAVCHIK, Anibal; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). MILSHINA, Natalia; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). WEI, Ming-Hui; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). KETCHUM, Karen, A.; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). DI FRANCESCO, Valentina; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). BEASLEY, Ellen, M.; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US).

(74) Agent: CELERA GENOMICS; Millman, Robert, A., Chief Intellectual Property Counsel, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,

[Continued on next page]

(54) Title: ISOLATED HUMAN G-PROTEIN COUPLED RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR PROTEINS, AND USES THEREOF

(57) Abstract: The present invention provides amino acid sequences of peptides that are encoded by genes within the human genome, the GPCR peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthlogs and paralogs of the GPCR peptides and methods of identifying modulators of the GPCR peptides.

FORM 1:

```
      1  ATGACGTCCA CCTGCACCAA CAGCACGCGC GAGAGTAACA GCAGCCACAC
     51  GTGCATGCCC CTCTCCAAAA TGCCCATCAG CCTGGCCCAC GGCATCATCC
    101  GCTCAACGGT GCTGGTTATC TTCCTCGCCG CCTCTTTCGT CGGCAACATA
    151  GTGCTGGCGC TAGTGTTGCA GCGCAAGCCG CAGCTGCTGC AGGTGACCAA
    201  CCGTTTTATC TTTAACCTCC TCGTCACCGA CCTGCTGCAG ATTTCGCTCG
    251  TGGCCCCCTG GGTGGTGGCC ACCTCTGTGC CTCTCTTCTG GCCCCTCAAC
    301  AGCCACTTCT GCACGGCCCT GGTTAGCCTC ACCCACCTGT TCGCCTTCGC
    351  CACCGTCAAC ACCATTGTCG TGGTGTCAGT GCATCGCTAC TTGTCCATCA
    401  TCCACCCTCT CTCCTACCCG TCCAAGATGA CCCAGGGCCG CGGTTACCTG
    451  CTCCTCTATG GCACCTGGAT TGTGGCCATC CTGCAGAGCA CTCCTCCACT
    501  CTACGGCTGG GCCCAGGCTG CCTTTGATGA GCGCAATGCT CTCTGCTCCA
    551  TGATCTGGGG GGCCAGCCCC AGCTACACTA TTCTCAGCGT GGTGTCCTTC
    601  ATCGTCATTC CACTGATTGT CATGATTGCC TGCTACTCCG TGGTGTTCTG
    651  TGCAGCCCGG AGGCAGCATG CTCTGCTGTA CAATGTCAAG AGACACAGCT
    701  TGGAAGTGCG AGTCAAGGAC TGTGTGGAGA ATGAGGATGA AGAGGGAGCA
    751  GAGAAGAAGG AGGAGTTCCA GGATGAGACT GAGTTTCGCC GCCAGCATGA
    801  AGGTGAGGTC AAGGCCAAGG AGGGCAGAAT GGAAGCCAAG GACGGCAGCC
    851  TGAAGCCCAA GGAAGGAAGC ACGGGGACCA GTGAGAGTAG TGTAGAGGCC
    901  ACGGGCAGCG AGGAGGTCAG AGAGAGCAGC ACGGTGGCCA GCGACGCCAG
    951  CATGGACGGT AAGGAACGCA GCACCAAAGT TCAGCAGAAC AGCATGAAGG
   1001  CAGACAAGGG TCGCACAGAG GTCAACCAGT GCAGCATTGA CTTGGGTGAA
   1051  GATGACATGG AGTTTGGTGA AGACGACATC AATTTCAGTG AGGATGACGT
   1101  CGAGGCAGTG AACATCCCGG AGAGCCTCCC ACCCAGTCGT CGTAACAGCA
   1151  ACAGCAACCC TCCTCTGCCC AGGTGCTACC AGTGCAAAGC TGCTAAAGTG
   1201  ATCTTCATCA TCATTTTC?C CTATGTGCTA TCCCTGGGGC CCTACTGCTT
   1251  TTTAGCAGTC CTGGCCGTGT GGGTGGATGT CGAAACCCAG GTACCCCAGT
   1301  GGGTGATCAC CATAATCATC TGGCTTTTCT TCCTGCAGTG CTGCATCCAC
   1351  CCCTATGTCT ATGGCTACAT GCACAAGACC ATTAAGAAGG AATCCAGGA
   1401  CATGCTGAAG AAGTTCTTCT GCAAGGAAAA GCCCCCGAAA GAAGATAGCC
   1451  ACCCAGACCT GCCCGGAACA GAGGGTGGGA CTGAAGGCAA GATTGTCCCT
   1501  TCCTACGATT CTGCTACTTT TCCTTGA
         (SEQ ID NO:1)
```

FEATURES:
Start: 1
Stop: 1525

FORM 2:

```
      1 ATGACGTCCA CCTGCACCAA CAGCACGCGC GAGAGTAACA GCAGCCACAC
     51 GTGCATGCCC CTCTCCAAAA TGCCCATCAG CCTGGCCCAC GGCATCATCC
    101 GCTCAACGGT GCTGGTTATC TTCCTCGCCG CCTCTTTCGT CGGCAACATA
    151 GTGCTGGCGC TAGTGTTGCA GCGCAAGCCG CAGCTGCTGC AGGTGACCAA
    201 CCGTTTTATC TTTAACCTCC TCGTCACCGA CCTGCTGCAG ATTTCGCTCG
    251 TGGCCCCCTG GGTGGTGGCC ACCTCTGTGC CTCTCTTCTG GCCCCTCAAC
    301 AGCCACTTCT GCACGGCCCT GGTTAGCCTC ACCCACCTGT TCGCCTTCGC
    351 CACCGTCAAC ACCATTGTCG TGGTGTCAGT GCATCGCTAC TTGTCCATCA
    401 TCCACCCTCT CTCCTACCCG TCCAAGATGA CCCAGGGCCG CGGTTACCTG
    451 CTCCTCTATG GCACCTGGAT TGTGGCCATC CTGCAGAGCA CTCCTCCACT
    501 CTACGGCTGG GCCCAGGCTG CCTTTGATGA GCGCAATGCT CTCTGCTCCA
    551 TGATCTGGGG GGCCAGCCCC AGCTACACTA TTCTCAGCGT GGTGTCCTTC
    601 ATCGTCATTC CACTGATTGT CATGATTGCC TGCTACTCCG TGGTGTTCTG
    651 TGCAGCCCGG AGGCAGCATG CTCTGCTGTA CAATGTCAAG AGACACAGCT
    701 TGGAAGTGCG AGTCAAGGAC TGTGTGGAGA ATGAGGATGA AGAGGGAGCA
    751 GAGAAGAAGG AGGAGTTCCA GGATGAGACT GAGTTTCGCC GCCAGCATGA
    801 AGGTGAGGTC AAGGCCAAGG AGGGCAGAAT GGAAGCCAAG GACGGCAGCC
    851 TGAAGCCCAA GGAAGGAAGC ACGGGGACCA GTGAGAGTAG TGTAGAGGCC
    901 ACGGGCAGCG AGGAGGTCAG AGAGAGCAGC ACGGTGGCCA GCGACGCCAG
    951 CATGGACGGT AAGGAACGCA GCACCAAAGT TCAGCAGAAC AGCATGAAGG
   1001 CAGACAAGGG TCGCACAGAG GTCAACCAGT GCAGCATTGA CTTGGGTGAA
   1051 GATGACATGG AGTTTGGTGA AGACGACATC AATTTCAGTG AGGATGACGT
   1101 CGAGGCAGTG AACATCCCGG AGAGCCTCCC ACCCAGTCGT CGTAACAGCA
   1151 ACAGCAACCC TCCTCTGCCC AGGTGCTACC AGTGCAAAGC TGCTAAAGTG
   1201 ATCTTCATCA TCATTTTCTC CTATGTGCTA TCCCTGGGGC CCTACTGCTT
   1251 TTTAGCAGTC CTGGCCGTGT GGGTGGATGT CGAAACCCAG GTACCCCAGT
   1301 GGGTGATCAC CATAATCATC TGGCTTTTCT TCCTGCAGTG CTGCATCCAC
   1351 CCCTATGTCT ATGGCTACAT GCACAAGACC ATTAAGAAGG AATCCAGGA
   1401 CATGCTGAAG AAGTTCTTCT GCAAGGAAAA GCCCCCGAAA GAAGATAGCC
   1451 ACCCAGACCT GCCCGGAACA GAGGGTGGGA CTGAAGGCAA GATTGTCCCT
   1501 TCCTACGATT CTGCTACTTT TCCTTGA
        (SEQ ID NO:4)
```

FEATURES:
Start: 1
Stop: 1195

GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

ISOLATED HUMAN G-PROTEIN COUPLED RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR PROTEINS, AND USES THEREOF

RELATED APPLICATIONS

The present application is a Continuation-In-Part of U.S. Serial No.09/638,018, filed on August 14, 2000 (Atty. Docket CL000754) and Continuation-In-Part of U.S. Serial No. 09/769,741, filed on January 26, 2001 (Atty. Docket CL000754CIP).

FIELD OF THE INVENTION

The present invention is in the field of G-Protein coupled receptors (GPCRs) that are related to the adrenergic receptor subfamily, recombinant DNA molecules, and protein production. The present invention specifically provides novel GPCR peptides and proteins and nucleic acid molecules encoding such peptide and protein molecules, all of which are useful in the development of human therapeutics and diagnostic compositions and methods.

BACKGROUND OF THE INVENTION

G-protein coupled receptors

G-protein coupled receptors (GPCRs) constitute a major class of proteins responsible for transducing a signal within a cell. GPCRs have three structural domains: an amino terminal extracellular domain, a transmembrane domain containing seven transmembrane segments, three extracellular loops, and three intracellular loops, and a carboxy terminal intracellular domain. Upon binding of a ligand to an extracellular portion of a GPCR, a signal is transduced within the cell that results in a change in a biological or physiological property of the cell. GPCRs, along with G-proteins and effecters (intracellular enzymes and channels modulated by G-proteins), are the components of a modular signaling system that connects the state of intracellular second messengers to extracellular inputs.

GPCR genes and gene-products are potential causative agents of disease (Spiegel *et al., J. Clin. Invest. 92*:1119-1125 (1993); McKusick *et al., J. Med. Genet. 30*:1-26 (1993)). Specific defects in the rhodopsin gene and the V2 vasopressin receptor gene have been shown to cause various forms of retinitis pigmentosum (Nathans *et al., Annu. Rev. Genet. 26*:403-424(1992)), and

nephrogenic diabetes insipidus (Holtzman *et al.*, *Hum. Mol. Genet. 2*:1201-1204 (1993)). These receptors are of critical importance to both the central nervous system and peripheral physiological processes. Evolutionary analyses suggest that the ancestor of these proteins originally developed in concert with complex body plans and nervous systems.

The GPCR protein superfamily can be divided into five families: Family I, receptors typified by rhodopsin and the β2-purinergic receptor and currently represented by over 200 unique members (Dohlman *et al.*, *Annu. Rev. Biochem. 60*:653-688 (1991)); Family II, the parathyroid hormone/calcitonin/secretin receptor family (Juppner *et al.*, *Science 254*:1024-1026 (1991); Lin *et al.*, *Science 254*:1022-1024 (1991)); Family III, the metabotropic glutamate receptor family (Nakanishi, *Science 258* 597:603 (1992)); Family IV, the cAMP receptor family, important in the chemotaxis and development of *D. discoideum* (Klein *et al.*, *Science 241*:1467-1472 (1988)); and Family V, the fungal mating pheromone receptors such as STE2 (Kurjan, *Annu. Rev. Biochem. 61*:1097-1129 (1992)).

There are also a small number of other proteins that present seven putative hydrophobic segments and appear to be unrelated to GPCRs; they have not been shown to couple to G-proteins. *Drosophila* expresses a photoreceptor-specific protein, bride of sevenless (boss), a seven-transmembrane-segment protein that has been extensively studied and does not show evidence of being a GPCR (Hart *et al.*, *Proc. Natl. Acad. Sci. USA 90*:5047-5051 (1993)). The gene *frizzled (fz)* in *Drosophila* is also thought to be a protein with seven transmembrane segments. Like boss, fz has not been shown to couple to G-proteins (Vinson *et al.*, *Nature 338*:263-264 (1989)).

G proteins represent a family of heterotrimeric proteins composed of α, β and γ subunits, that bind guanine nucleotides. These proteins are usually linked to cell surface receptors, e.g., receptors containing seven transmembrane segments. Following ligand binding to the GPCR, a conformational change is transmitted to the G protein, which causes the α-subunit to exchange a bound GDP molecule for a GTP molecule and to dissociate from the βγ-subunits. The GTP-bound form of the α-subunit typically functions as an effector-modulating moiety, leading to the production of second messengers, such as cAMP (e.g., by activation of adenyl cyclase), diacylglycerol or inositol phosphates. Greater than 20 different types of α-subunits are known in humans. These subunits associate with a smaller pool of β and γ subunits. Examples of mammalian G proteins include Gi, Go, Gq, Gs and Gt. G proteins are described extensively in Lodish *et al.*, *Molecular Cell Biology*, (Scientific American Books Inc., New York, N.Y., 1995), the contents of which are incorporated herein by reference. GPCRs, G proteins and G protein-linked effector and second messenger systems have been reviewed in *The G-Protein Linked Receptor Fact*

Book, Watson *et al.*, eds., Academic Press (1994).

Aminergic GPCRs

One family of the GPCRS, Family II, contains receptors for acetylcholine, catecholamine, and indoleamine ligands (hereafter referred to as biogenic amines). The biogenic amine receptors (aminergic GPCRs) represent a large group of GPCRs that share a common evolutionary ancestor and which are present in both vertebrate (deuterostome), and invertebrate (protostome) lineages. This family of GPCRs includes, but is not limited to the 5-HT-like, the dopamine-like, the acetylcholine-like, the adrenaline-like and the melatonin-like GPCRs.

Dopamine receptors

The understanding of the dopaminergic system relevance in brain function and disease developed several decades ago from three diverse observations following drug treatments. These were the observations that dopamine replacement therapy improved Parkinson's disease symptoms, depletion of dopamine and other catecholamines by reserpine caused depression and antipsychotic drugs blocked dopamine receptors. The finding that the dopamine receptor binding affinities of typical antipsychotic drugs correlate with their clinical potency led to the dopamine overactivity hypothesis of schizophrenia (Snyder, S.H., *Am J Psychiatry 133*, 197-202 (1976); Seeman, P. and Lee, T., *Science 188*, 1217-9 (1975)). Today, dopamine receptors are crucial targets in the pharmacological therapy of schizophrenia, Parkinson's disease, Tourette's syndrome, tardive dyskinesia and Huntington's disease. The dopaminergic system includes the nigrostriatal, mesocorticolimbic and tuberoinfundibular pathways. The nigrostriatal pathway is part of the striatal motor system and its degeneration leads to Parkinson's disease; the mesocorticolimbic pathway plays a key role in reinforcement and in emotional expression and is the desired site of action of antipsychotic drugs; the tuberoinfundibular pathways regulates prolactin secretion from the pituitary.

Dopamine receptors are members of the G protein coupled receptor superfamily, a large group proteins that share a seven helical membrane-spanning structure and transduce signals through coupling to heterotrimeric guanine nucleotide-binding regulatory proteins (G proteins). Dopamine receptors are classified into subfamilies: D1-like (D1 and D5) and D2-like (D2, D3 and D4) based on their different ligand binding profiles, signal transduction properties, sequence homologies and genomic organizations (Civelli, O., Bunzow, J.R. and Grandy, D.K., *Annu Rev Pharmacol Toxicol 33*, 281-307 (1993)). The D1-like receptors, D1 and D5, stimulate cAMP synthesis through coupling with Gs-like proteins and their genes do not contain introns within their

protein coding regions. On the other hand, the D2-like receptors, D2, D3 and D4, inhibit cAMP synthesis through their interaction with Gi-like proteins and share a similar genomic organization which includes introns within their protein coding regions.

Serotonin receptors

Serotonin (5-Hydroxytryptamine; 5-HT) was first isolated from blood serum, where it was shown to promote vasoconstriction (Rapport, M.M., Green, A.A. and Page, I.H., *J Biol Chem 176*, 1243-1251 (1948). Interest on a possible relationship between 5-HT and psychiatric disease was spurred by the observations that hallucinogens such as LSD and psilocybin inhibit the actions of 5-HT on smooth muscle preparations (Gaddum, J.H. and Hameed, K.A., *Br J Pharmacol 9*, 240-248 (1954)). This observation lead to the hypothesis that brain 5-HT activity might be altered in psychiatric disorders (Wooley, D.W. and Shaw, E., *Proc Natl Acad Sci U S A 40*, 228-231 (1954); Gaddum, J.H. and Picarelli, Z.P., *Br J Pharmacol 12*, 323-328 (1957)). This hypothesis was strengthened by the introduction of tricyclic antidepressants and monoamine oxidase inhibitors for the treatment of major depression and the observation that those drugs affected noradrenaline and 5-HT metabolism. Today, drugs acting on the serotoninergic system have been proved to be effective in the pharmacotherapy of psychiatric diseases such as depression, schizophrenia, obsessive-compulsive disorder, panic disorder, generalized anxiety disorder and social phobia as well as migraine, vomiting induced by cancer chemotherapy and gastric motility disorders.

Serotonin receptors represent a very large and diverse family of neurotransmitter receptors. To date thirteen 5-HT receptor proteins coupled to G proteins plus one ligand–gated ion channel receptor (5-HT3) have been described in mammals. This receptor diversity is thought to reflect serotonin's ancient origin as a neurotransmitter and a hormone as well as the many different roles of 5-HT in mammals. The 5-HT receptors have been classified into seven subfamilies or groups according to their different ligand-binding affinity profiles, molecular structure and intracellular transduction mechanisms (Hoyer, D. et al., *Pharmacol. Rev. 46*, 157-203 (1994)).

Adrenergic GPCRs

The present invention provides two possible isoforms of a novel human GPCR, transcripts corresponding to each possible isoform, the genomic sequence for the gene that encodes either isoform, and possible gene structures corresponding to each of the two isoforms. The two GPCR isoforms are herein referred to as "form 1" and "form 2". Form 1 has been previously disclosed by applicant in U.S. application 09/638,018, filed August 14, 2000. Specifically, form 1 comprises a single exon structure whereas form 2 comprises two exons with

intervening intronic DNA. These structural differences are indicated in the genomic features provided in Figure 3 and in the amino acid sequence alignment of form 1 and form 2 provided in Figure 2. Additionally, the transcript sequence of each form is provided in Figure 1 and the protein/amino acid sequence of each form is provided in Figure 2.

The adrenergic receptors comprise one of the largest and most extensively characterized families within the G-protein coupled receptor "superfamily". This superfamily includes not only adrenergic receptors, but also muscarinic, cholinergic, dopaminergic, serotonergic, and histaminergic receptors. Numerous peptide receptors include glucagon, somatostatin, and vasopressin receptors, as well as sensory receptors for vision (rhodopsin), taste, and olfaction, also belong to this growing family. Despite the diversity of signalling molecules, G-protein coupled receptors all possess a similar overall primary structure, characterized by 7 putative membrane-spanning .alpha. helices (Probst et al., 1992). In the most basic sense, the adrenergic receptors are the physiological sites of action of the catecholamines, epinephrine and norepinephrine. Adrenergic receptors were initially classified as either .alpha. or .beta. by Ahlquist, who demonstrated that the order of potency for a series of agonists to evoke a physiological response was distinctly different at the 2 receptor subtypes (Ahlquist, 1948). Functionally, .alpha. adrenergic receptors were shown to control vasoconstriction, pupil dilation and uterine inhibition, while .beta. adrenergic receptors were implicated in vasorelaxation, myocardial stimulation and bronchodilation (Regan et al., 1990). Eventually, pharmacologists realized that these responses resulted from activation of several distinct adrenergic receptor subtypes. .beta. adrenergic receptors in the heart were defined as .beta..sub.1, while those in the lung and vasculature were termed .beta..sub.2 (Lands et al., 1967).

.alpha. Adrenergic receptors, meanwhile, were first classified based on their anatomical location, as either pre or post-synaptic (.alpha..sub.2 and .alpha..sub.1, respectively) (Langer et al., 1974). This classification scheme was confounded, however, by the presence of .alpha..sub.2 receptors in distinctly non-synaptic locations, such as platelets (Berthelsen and Pettinger, 1977). With the development of radioligand binding techniques, .alpha. adrenergic receptors could be distinguished pharmacologically based on their affinities for the antagonists prazosin or yohimbine (Stark, 1981). Definitive evidence for adrenergic receptor subtypes, however, awaited purification and molecular cloning of adrenergic receptor subtypes. In 1986, the genes for the hamster .beta..sub.2 (Dickson et al., 1986) and turkey .beta..sub.1 adrenergic receptors (Yarden et al., 1986) were cloned and sequenced. Hydropathy analysis revealed that these proteins contain 7 hydrophobic domains similar to rhodopsin, the receptor for light. Since that time the

adrenergic receptor family has expanded to include 3 subtypes of .beta. receptors (Emorine et al., 1989), 3 subtypes of .alpha..sub.1 receptors (Schwinn et al., 1990), and 3 distinct types of .beta..sub.2 receptors (Lomasney et al., 1990).

The cloning, sequencing and expression of alpha receptor subtypes from animal tissues has led to the subclassification of the alpha 1 receptors into alpha 1d (formerly known as alpha 1a or 1a/1d), alpha 1b and alpha 1a (formerly known as alpha 1c) subtypes. Each alpha 1 receptor subtype exhibits its own pharmacologic and tissue specificities. The designation "alpha 1a" is the appellation recently approved by the IUPHAR Nomenclature Committee for the previously designated "alpha 1c" cloned subtype as outlined in the 1995 Receptor and Ion Channel Nomenclature Supplement (Watson and Girdlestone, 1995). The designation alpha 1a is used throughout this application to refer to this subtype. At the same time, the receptor formerly designated alpha 1a was renamed alpha 1d. The new nomenclature is used throughout this application. Stable cell lines expressing these alpha 1 receptor subtypes are referred to herein; however, these cell lines were deposited with the American Type Culture Collection (ATCC) under the old nomenclature. For a review of the classification of alpha 1 adrenoceptor subtypes, see, Martin C. Michel, et al., Naunyn-Schmiedeberg's Arch. Pharmacol. (1995) 352:1-10.

The differences in the alpha adrenergic receptor subtypes have relevance in pathophysiologic conditions. Benign prostatic hyperplasia, also known as benign prostatic hypertrophy or BPH, is an illness typically affecting men over fifty years of age, increasing in severity with increasing age. The symptoms of the condition include, but are not limited to, increased difficulty in urination and sexual dysfunction. These symptoms are induced by enlargement, or hyperplasia, of the prostate gland. As the prostate increases in size, it impinges on free-flow of fluids through the male urethra. Concommitantly, the increased noradrenergic innervation of the enlarged prostate leads to an increased adrenergic tone of the bladder neck and urethra, further restricting the flow of urine through the urethra.

The .alpha..sub.2 receptors appear to have diverged rather early from either .beta. or .alpha..sub.1 receptors. The .alpha..sub.2 receptors have been broken down into 3 molecularly distinct subtypes termed .alpha..sub.2 C2, .alpha..sub.2 C4, and .alpha..sub.2 C10 based on their chromosomal location. These subtypes appear to correspond to the pharmacologically defined .alpha..sub.2B, .alpha..sub.2C, and .alpha..sub.2A subtypes, respectively (Bylund et al., 1992). While all the receptors of the adrenergic type are recognized by epinephrine, they are pharmacologically distinct and are encoded by separate genes. These receptors are generally coupled to different second messenger pathways that are linked through G-proteins. Among the

adrenergic receptors, .beta..sub.1 and .beta..sub.2 receptors activate the adenylate cyclase, .alpha..sub.2 receptors inhibit adenylate cyclase and .alpha..sub.1 receptors activate phospholipase C pathways, stimulating breakdown of polyphosphoinositides (Chung, F. Z. et al., J. Biol. Chem.,' 263:4052 (1988)). .alpha..sub.1 and .alpha..sub.2 adrenergic receptors differ in their cell activity for drugs.

Issued US patent that disclose the utility of members of this family of proteins include, but are not limited to, 6,063,785 Phthalimido arylpiperazines useful in the treatment of benign prostatic hyperplasia; 6,060,492 Selective .beta.3 adrenergic agonists; 6,057,350 Alpha 1a adrenergic receptor antagonists; 6,046,192 Phenylethanolaminotetralincarboxamide derivatives; 6,046,183 Method of synergistic treatment for benign prostatic hyperplasia; 6,043,253 Fused piperidine substituted arylsulfonamides as .beta.3-agonists; 6,043,224 Compositions and methods for treatment of neurological disorders and neurodegenerative diseases; 6,037,354 Alpha 1a adrenergic receptor antagonists; 6,034,106 Oxadiazole benzenesulfonamides as selective .beta..sub.3 Agonist for the treatment of Diabetes and Obesity; 6,011,048 Thiazole benzenesulfonamides as .beta.3 agonists for treatment of diabetes and obesity; 6,008,361 5,994,506 Adrenergic receptor; 5,994,294 Nitrosated and nitrosylated .alpha.-adrenergic receptor antagonist compounds, compositions and their uses; 5,990,128 .alpha..sub.1C specific compounds to treat benign prostatic hyperplasia; 5,977,154 Selective .beta.3 adrenergic agonist; 5,977,115 Alpha 1a adrenergic receptor antagonists; 5,939,443 Selective .beta.3 adrenergic agonists; 5,932,538 Nitrosated and nitrosylated .alpha.-adrenergic receptor antagonist compounds, compositions and their uses; 5,922,722 Alpha 1a adrenergic receptor antagonists 26 5,908,830 and 5,861,309 DNA endoding human alpha 1 adrenergic receptors.

GPCRs, particularly members of the adrenergic receptor subfamily, are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown GPCRs. The present invention advances the state of the art by providing a previously unidentified human GPCR.

SUMMARY OF THE INVENTION

The present invention is based in part on the identification of nucleic acid sequences that encode amino acid sequences of human GPCR peptides and proteins that are related to the adrenergic subfamily, allelic variants thereof and other mammalian orthologs thereof. These

unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents.

The proteins of the present inventions are GPCRs that participate in signaling pathways mediated by the adrenergic subfamily in cells that express these proteins. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain. As used herein, a "signaling pathway" refers to the modulation (e.g., stimulation or inhibition) of a cellular function/activity upon the binding of a ligand to the GPCR protein. Examples of such functions include mobilization of intracellular molecules that participate in a signal transduction pathway, e.g., phosphatidylinositol 4,5-bisphosphate (PIP_2), inositol 1,4,5-triphosphate (IP_3) and adenylate cyclase; polarization of the plasma membrane; production or secretion of molecules; alteration in the structure of a cellular component; cell proliferation, e.g., synthesis of DNA; cell migration; cell differentiation; and cell survival

The response mediated by the receptor protein depends on the type of cell it is expressed on. Some information regarding the types of cells that express other members of the subfamily of GPCRs of the present invention is already known in the art (see references cited in Background and information regarding closest homologous protein provided in Figure 2; Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain.). For example, in some cells, binding of a ligand to the receptor protein may stimulate an activity such as release of compounds, gating of a channel, cellular adhesion, migration, differentiation, etc., through phosphatidylinositol or cyclic AMP metabolism and turnover while in other cells, the binding of the ligand will produce a different result. Regardless of the cellular activity/response modulated by the particular GPCR of the present invention, a skilled artisan will clearly know that the receptor protein is a GPCR and interacts with G proteins to produce one or more secondary signals, in a variety of intracellular signal transduction pathways, e.g., through phosphatidylinositol or cyclic AMP metabolism and turnover, in a cell thus participating in a biological process in the cells or tissues that express the GPCR. Experimental data as provided in Figure 1 indicates expression in the brain as determined by cDNA retrieval.

As used herein, "phosphatidylinositol turnover and metabolism" refers to the molecules involved in the turnover and metabolism of phosphatidylinositol 4,5-bisphosphate (PIP_2) as well as to the activities of these molecules. PIP_2 is a phospholipid found in the cytosolic leaflet of the plasma membrane. Binding of ligand to the receptor activates, in some cells, the plasma-membrane enzyme phospholipase C that in turn can hydrolyze PIP_2 to produce 1,2-diacylglycerol (DAG) and

inositol 1,4,5-triphosphate (IP_3). Once formed IP_3 can diffuse to the endoplasmic reticulum surface where it can bind an IP_3 receptor, e.g., a calcium channel protein containing an IP_3 binding site. IP_3 binding can induce opening of the channel, allowing calcium ions to be released into the cytoplasm. IP_3 can also be phosphorylated by a specific kinase to form inositol 1,3,4,5-tetraphosphate (IP_4), a molecule that can cause calcium entry into the cytoplasm from the extracellular medium. IP_3 and IP_4 can subsequently be hydrolyzed very rapidly to the inactive products inositol 1,4-biphosphate (IP_2) and inositol 1,3,4-triphosphate, respectively. These inactive products can be recycled by the cell to synthesize PIP_2. The other second messenger produced by the hydrolysis of PIP_2, namely 1,2-diacylglycerol (DAG), remains in the cell membrane where it can serve to activate the enzyme protein kinase C. Protein kinase C is usually found soluble in the cytoplasm of the cell, but upon an increase in the intracellular calcium concentration, this enzyme can move to the plasma membrane where it can be activated by DAG. The activation of protein kinase C in different cells results in various cellular responses such as the phosphorylation of glycogen synthase, or the phosphorylation of various transcription factors, e.g., NF-kB. The language "phosphatidylinositol activity", as used herein, refers to an activity of PIP_2 or one of its metabolites.

Another signaling pathway in which the receptor may participate is the cAMP turnover pathway. As used herein, "cyclic AMP turnover and metabolism" refers to the molecules involved in the turnover and metabolism of cyclic AMP (cAMP) as well as to the activities of these molecules. Cyclic AMP is a second messenger produced in response to ligand-induced stimulation of certain G protein coupled receptors. In the cAMP signaling pathway, binding of a ligand to a GPCR can lead to the activation of the enzyme adenyl cyclase, which catalyzes the synthesis of cAMP. The newly synthesized cAMP can in turn activate a cAMP-dependent protein kinase. This activated kinase can phosphorylate a voltage-gated potassium channel protein, or an associated protein, and lead to the inability of the potassium channel to open during an action potential. The inability of the potassium channel to open results in a decrease in the outward flow of potassium, which normally repolarizes the membrane of a neuron, leading to prolonged membrane depolarization.

By targeting an agent to modulate a GPCR, the signaling activity and biological process mediated by the receptor can be agonized or antagonized in specific cells and tissues. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain. Such agonism and antagonism serves as a basis for modulating a biological activity in a therapeutic context (mammalian therapy) or toxic context (anti-cell therapy, e.g. anti-cancer agent).

DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequence of the transcripts (form 1 = SEQ ID NO:1, form 2 = SEQ ID NO:4) that encode the GPCR of the present invention. In addition, structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of the inventions based on these molecular sequences. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain.

FIGURE 2 provides the predicted amino acid sequences (form 1 = SEQ ID NO:2, form 2 = SEQ ID NO:5) of the GPCR of the present invention. In addition, structure and functional information such as protein family, function, and modification sites is provided where available, allowing one to readily determine specific uses of the inventions based on these molecular sequences. Figure 2 also provides an amino acid sequence alignment of form 1 and form 2. For instances in the figures in which the form (i.e., form 1 or form 2) is not specified, any amino acid coordinates generally are based on the amino acid sequence provided for form 1.

FIGURE 3 provides the genomic sequence (SEQ ID NO:3) that spans the gene encoding the GPCR proteins of the present invention.. In addition, structure and functional information, such as intron/exon structure, promoter location, etc., is provided for form 1 and form 2, allowing one to readily determine specific uses of the inventions based on this molecular sequence. As illustrated in Figure 3, a non-synonymous coding SNP (T4126C) was identified in the genomic sequence. The chromosome map position is also provided in Figure 3.

DETAILED DESCRIPTION OF THE INVENTION

General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed previously unidentified fragments of the human genome that encode peptides that share structural and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a GPCR protein or part of a GPCR protein, that are related to the adrenergic subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis,

the present invention provides amino acid sequences of human GPCR peptides and proteins that are related to the adrenergic subfamily, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these GPCR peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the GPCR of the present invention.

In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known GPCR proteins of the adrenergic subfamily and the expression pattern observed. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain. The art has clearly established the commercial importance of members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the known adrenergic family or subfamily of GPCR proteins.

Specific Embodiments

Peptide Molecules

The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the GPCR family of proteins and are related to the adrenergic subfamily (protein sequences are provided in Figure 2, transcript/cDNA sequences are provided in Figure 1 and genomic sequences are provided in Figure 3). The peptide sequences provided in Figure 2, as well as the obvious variants described herein, particularly allelic variants as identified herein and using the information in Figure 3, will be referred herein as the GPCR peptides of the present invention, GPCR peptides, or peptides/proteins of the present invention.

The present invention provides isolated peptide and protein molecules that consist of, consist essentially of, or comprise the amino acid sequences of the GPCR peptides disclosed in Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA sequence, or Figure 3, genomic sequence), as well as all obvious variants of these peptides that are within

the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the GPCR peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

The isolated GPCR peptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain. For example, a nucleic acid molecule encoding the GPCR peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NOS:2 AND 5), for example, proteins encoded by the transcript nucleic acid sequences shown in Figure 1 (SEQ ID NOS:1 AND 4) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). The amino acid sequence of such a protein is provided in Figure 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NOS:2 AND 5), for example, proteins encoded by the transcript sequences shown in Figure 1 (SEQ ID NOS:1 AND 4) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that comprise the amino acid sequences provided in Figure 2 (SEQ ID NOS:2 AND 5), for example, proteins encoded by the transcript nucleic acid sequences shown in Figure 1 (SEQ ID NOS:1 AND 4) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the GPCR peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The GPCR peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a GPCR peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the GPCR peptide. "Operatively linked" indicates that the GPCR peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the GPCR peptide.

In some uses, the fusion protein does not affect the activity of the GPCR peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant GPCR peptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-

frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A GPCR peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the GPCR peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the GPCR peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two

sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al., Nucleic Acids Res. 12(1)*:387 (1984)) (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST

programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the GPCR peptides of the present invention as well as being encoded by the same genetic locus as the GPCR peptide provided herein. The gene encoding the novel GPCR proteins of the present invention is located on a genome component that has been mapped to human chromosome X (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

Allelic variants of a GPCR peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the GPCR peptide as well as being encoded by the same genetic locus as the GPCR peptide provided herein. Genetic locus can readily be determined based on the genomic information provided in Figure 3, such as the genomic sequence mapped to the reference human. The gene encoding the novel GPCR proteins of the present invention is located on a genome component that has been mapped to human chromosome X (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data. As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Figure 3 provides information on a non-synonymous coding SNP that has been found in the gene encoding the GPCR proteins of the present invention. Specifically, a T/C polymorphism was identified at position 4126. The change in the amino acid sequence caused by this SNP is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequences provided in Figure 2 as a reference.

Paralogs of a GPCR peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the GPCR peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide

encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a GPCR peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the GPCR peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the GPCR peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the GPCR peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a GPCR peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science 247*:1306-1310 (1990).

Variant GPCR peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind ligand, ability to bind G-protein, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 2 provides the result of protein analysis that identifies critical domains/regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science 244*:1081-1085 (1989)), particularly using the results provided in Figure 2. The latter procedure

introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as ligand/effector molecule binding or in assays such as an *in vitro* proliferative activity. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al., J. Mol. Biol. 224*:899-904 (1992); de Vos *et al. Science 255*:306-312 (1992)).

The present invention further provides fragments of the GPCR peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly those comprising the residues identified in Figure 2. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16, or more contiguous amino acid residues from a GPCR peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the GPCR peptide or could be chosen for the ability to perform a function, e.g. ability to bind ligand or effector molecule or act as an immunogen. Particularly important fragments are biologically active fragments, peptides which are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the GPCR peptide, e.g., active site, a G-protein binding site, a transmembrane domain or a ligand-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well-known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in Figure 2.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in GPCR peptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art(some of these features are identified in Figure 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid

derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well-known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol. 182*: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci. 663*:48-62 (1992)).

Accordingly, the GPCR peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature GPCR peptide is fused with another compound, such as a compound to increase the half-life of the GPCR peptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature GPCR peptide, such as a leader or secretory sequence or a sequence for purification of the mature GPCR peptide or a pro-protein sequence.

Protein/Peptide Uses

The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures and Back Ground Section; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or receptor) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the binding partner so as to develop a system to identify inhibitors of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as commercial products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, GPCRs isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug, particularly in modulating a biological or pathological response in a cell or tissue that expresses the GPCR. Experimental data as provided in Figure 1 indicates expression in the brain as determined by cDNA retrieval. Approximately 70% of all pharmaceutical agents modulate the activity of a GPCR. A combination of the invertebrate and mammalian ortholog can be used in selective screening methods to find agents specific for invertebrates. The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention, particularly in combination with the expression information provided in Figure 1. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain. Such uses can readily be determined using the information provided herein, that known in the art and routine experimentation.

The proteins of the present invention (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to GPCRs that are related to members of the adrenergic subfamily. Such assays involve any of the known GPCR functions or activities or properties useful for diagnosis and treatment of GPCR-related conditions that are specific for the subfamily of GPCRs that the one of the present invention belongs to, particularly in cells and tissues that express this receptor. Experimental data as provided in Figure 1 indicates expression in the brain as determined by cDNA retrieval.

The proteins of the present invention are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the receptor protein, as a biopsy or expanded in cell culture. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the receptor protein.

The polypeptides can be used to identify compounds that modulate receptor activity of the protein in its natural state, or an altered form that causes a specific disease or pathology associated with the receptor. Both the GPCRs of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the receptor. These compounds can be further screened against a functional receptor to determine the effect of the compound on the receptor activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the receptor to a desired degree.

Further, the proteins of the present invention can be used to screen a compound for the ability to stimulate or inhibit interaction between the receptor protein and a molecule that normally interacts with the receptor protein, e.g. a ligand or a component of the signal pathway that the receptor protein normally interacts (for example, a G-protein or other interactor involved in cAMP or phosphatidylinositol turnover and/or adenylate cyclase, or phospholipase C activation). Such assays typically include the steps of combining the receptor protein with a candidate compound under conditions that allow the receptor protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the receptor protein and the target, such as any of the associated effects of signal transduction such as G-protein phosphorylation, cAMP or phosphatidylinositol turnover, and adenylate cyclase or phospholipase C activation.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature 354*:82-84 (1991); Houghten *et al.*, *Nature 354*:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell 72*:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, $F(ab')_2$, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the receptor that competes for ligand binding. Other candidate compounds include mutant receptors or appropriate fragments containing mutations that affect receptor function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) receptor activity. The assays typically involve an assay of events in the signal transduction pathway that indicate receptor activity. Thus, a cellular process such as proliferation, the expression of genes that are up- or down-regulated in response to the receptor protein dependent signal cascade, can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase.

Any of the biological or biochemical functions mediated by the receptor can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the Figures, particularly Figure 2. Specifically, a biological function of a cell or tissues that expresses the receptor can be assayed. Experimental data as provided in Figure 1 indicates expression in the brain as determined by cDNA retrieval.

Binding and/or activating compounds can also be screened by using chimeric receptor proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a G-protein-binding region can be used that interacts with a different G-protein then that which is recognized by the native receptor. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. Alternatively, the entire transmembrane portion or subregions (such as transmembrane segments or intracellular or extracellular loops) can be replaced with the entire transmembrane portion or subregions specific to a host cell that is different from the host cell from which the amino terminal extracellular domain and/or the G-protein-binding region are derived. This allows for assays to be performed in other than the specific host cell from which the receptor is derived. Alternatively, the amino terminal extracellular domain (and/or other ligand-binding regions) could be replaced by a domain (and/or other binding region) binding a different ligand, thus, providing an assay for test compounds that interact with the heterologous amino terminal extracellular domain (or region) but still cause signal transduction. Finally, activation can be detected by a reporter gene containing an easily detectable coding region operably linked to a transcriptional regulatory sequence that is part of the native signal transduction pathway.

The proteins of the present invention are also useful in competition binding assays in methods designed to discover compounds that interact with the receptor. Thus, a compound is

exposed to a receptor polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide (Hodgson, Bio/technology, 1992, Sept 10(9);973-80). Soluble receptor polypeptide is also added to the mixture. If the test compound interacts with the soluble receptor polypeptide, it decreases the amount of complex formed or activity from the receptor target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the receptor. Thus, the soluble polypeptide that competes with the target receptor region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the receptor protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ^{35}S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of receptor-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a receptor-binding protein and a candidate compound are incubated in the receptor protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the receptor protein target molecule, or which are reactive with receptor protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Agents that modulate one of the GPCRs of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

Modulators of receptor protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the receptor pathway, by treating cells or tissues that express the GPCR. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain. These methods of treatment include the steps of administering a modulator of the GPCR's activity in a pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described herein.

In yet another aspect of the invention, the GPCR proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with the GPCR and are involved in GPCR activity. Such GPCR-binding proteins are also likely to be involved in the propagation of signals by the GPCR proteins or GPCR targets as, for example, downstream elements of a GPCR-mediated signaling pathway. Alternatively, such GPCR-binding proteins are likely to be GPCR inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a GPCR protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming a GPCR-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the GPCR protein.

This invention further pertains to novel agents identified by the above-described

screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a GPCR modulating agent, an antisense GPCR nucleic acid molecule, a GPCR-specific antibody, or a GPCR-binding partner) can be used in an animal or other model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The GPCR proteins of the present invention are also useful to provide a target for diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain. The method involves contacting a biological sample with a compound capable of interacting with the receptor protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered receptor activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagent,

such as an antibody or protein binding agent. Alternatively, the peptide can be detected *in vivo* in a subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the receptor protein in which one or more of the receptor functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and receptor activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain. Accordingly, methods for treatment include the use of the GPCR protein or fragments.

Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or $F(ab')_2$, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, Antibodies, Cold Spring Harbor Press, (1989).

In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods and as presented in the Figures.

Antibodies are preferably prepared from regions or discrete fragments of the GPCR proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or receptor/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues.

Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2).

Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Experimental data as provided in Figure 1 indicates expression in the brain as determined by cDNA retrieval. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. Experimental data as provided in Figure 1 indicates that GPCR

proteins of the present invention are expressed in the brain. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the GPCR peptide to a binding partner such as a ligand. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information relating to the proteins of the present invention.

The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nucleic acid arrays and similar methods have been developed for

antibody arrays.

Nucleic Acid Molecules

The present invention further provides isolated nucleic acid molecules that encode a GPCR peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the GPCR peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NOS:1 AND 4, transcript sequences and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the proteins provided in

Figure 2, SEQ ID NOS:2 AND 5. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NOS:1 AND 4, transcript sequences and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the proteins provided in Figure 2, SEQ ID NOS:2 AND 5. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise the nucleotide sequences shown in Figure 1 or 3 (SEQ ID NOS:1 AND 4, transcript sequences and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the proteins provided in Figure 2, SEQ ID NOS:2 AND 5. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprises several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, human genomic sequences (Figure 3) and cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the Figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As

generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the GPCR peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the GPCR proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in Figures 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene termination sequences. Such fragments are useful in controlling heterologous gene expression and in developing screens to identify gene-modulating agents. A promoter can readily be identified as being 5' to the ATG start site in the genomic sequence provided in Figure 3.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more

nucleotides. Further, a fragment could at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene. The gene encoding the novel GPCR proteins of the present invention is located on a genome component that has been mapped to human chromosome X (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

Figure 3 provides information on a non-synonymous coding SNP that has been found in the gene encoding the GPCR proteins of the present invention. Specifically, a T/C polymorphism was identified at position 4126. The change in the amino acid sequence caused by this SNP is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequences provided in Figure 2 as a reference.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are

hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2. As illustrated in Figure 3, a non-synonymous coding SNP (T4126C) was identified in the genomic sequence.

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods. The gene encoding the novel GPCR proteins of the present invention is located on a genome component that has been mapped to human chromosome X (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data.

The nucleic acid molecules are also useful in making vectors containing the gene regulatory

regions of the nucleic acid molecules of the present invention.

The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Experimental data as provided in Figure 1 indicates expression in the brain as determined by cDNA retrieval. Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in GPCR protein expression relative to normal results.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a GPCR protein, such as by measuring a level of a receptor-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a receptor gene has been mutated. Experimental data as provided in Figure 1 indicates expression in the brain as determined by cDNA retrieval.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate GPCR nucleic acid expression.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the GPCR gene, particularly biological and pathological processes that are mediated by the GPCR in cells and tissues that express it. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain. The method typically includes assaying the ability of the compound to

modulate the expression of the GPCR nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired GPCR nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the GPCR nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

The assay for GPCR nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the GPCR protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of GPCR gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of GPCR mRNA in the presence of the candidate compound is compared to the level of expression of GPCR mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate GPCR nucleic acid expression, particularly to modulate activities within a cell or tissue that expresses the proteins. Experimental data as provided in Figure 1 indicates expression in the brain as determined by cDNA retrieval. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or nucleic acid expression.

Alternatively, a modulator for GPCR nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the GPCR nucleic acid expression in the cells and tissues that express the protein. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in the brain.

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating

compounds on the expression or activity of the GPCR gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in GPCR nucleic acid, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in GPCR genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally-occurring genetic mutations in the GPCR gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the GPCR gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a GPCR protein.

Individuals carrying mutations in the GPCR gene can be detected at the nucleic acid level by a variety of techniques. Figure 3 provides information on a non-synonymous coding SNP that has been found in the gene encoding the GPCR proteins of the present invention. Specifically, a T/C polymorphism was identified at position 4126. The change in the amino acid sequence caused by this SNP is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequences provided in Figure 2 as a reference. The gene encoding the novel GPCR proteins of the present invention is located on a genome component that has been mapped to human chromosome X (as indicated in Figure 3), which is supported by multiple lines of evidence, such as STS and BAC map data. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a

ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science 241*:1077-1080 (1988); and Nakazawa *et al.*, *PNAS 91*:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res. 23*:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a GPCR gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant GPCR gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C.W., (1995) *Biotechniques 19*:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr. 36*:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol. 38*:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science 230*:1242 (1985)); Cotton *et al.*, *PNAS 85*:4397 (1988); Saleeba *et al.*, *Meth. Enzymol. 217*:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS 86*:2766 (1989); Cotton *et al.*, *Mutat. Res. 285*:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl. 9*:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature 313*:495 (1985)). Examples of other techniques for detecting point mutations include selective oligonucleotide hybridization, selective

amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the GPCR gene in an individual in order to select an appropriate compound or dosage regimen for treatment. As illustrated in Figure 3, a non-synonymous coding SNP (T4126C) was identified in the genomic sequence.

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control GPCR gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of GPCR protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into GPCR protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of GPCR nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired GPCR nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the GPCR protein, such as ligand binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in GPCR gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired GPCR protein to treat the individual.

The invention also encompasses kits for detecting the presence of a GPCR nucleic acid in a biological sample. Experimental data as provided in Figure 1 indicates expression in the brain as determined by cDNA retrieval. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting GPCR nucleic acid in a biological sample;

means for determining the amount of GPCR nucleic acid in the sample; and means for comparing the amount of GPCR nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect GPCR protein mRNA or DNA.

Nucleic Acid Arrays

The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in Figures 1 and 3 (SEQ ID NOS:1, 3, and 4).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee et al., PCT application W095/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown et. al., US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a gene or genes of interest.

In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack

predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large scale correlation studies on the

sequences, expression patterns, mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of the GPCR proteins/peptides of the present invention. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes, at least one of which is a gene of the present invention and or alleles of the GPCR gene of the present invention. Figure 3 provides information on a non-synonymous coding SNP that has been found in the gene encoding the GPCR proteins of the present invention. Specifically, a T/C polymorphism was identified at position 4126. The change in the amino acid sequence caused by this SNP is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequences provided in Figure 2 as a reference.

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the Human genome disclosed herein. Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques,* Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al., Techniques in Immunocytochemistry,* Academic Press, Orlando, FL Vol. 1 (1 982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology,* Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the Human genome disclosed herein; and

(b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified GPCR genes of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

Vectors/host cells

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed

in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ, the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, eg. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold

Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene 67*:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene 69*:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology 185*:60-89 (1990)).

Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid

molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res. 20*:2111-2118 (1992)).

The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J. 6*:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell 30*:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene 54*:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol. 3*:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology 170*:31-39 (1989)).

In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature 329*:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J. 6*:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These

include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell- free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as GPCRs, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the peptide is not secreted into the medium, which is typically the case with GPCRs, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate

precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a GPCR protein or peptide that can be further purified to produce desired amounts of GPCR protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

Host cells are also useful for conducting cell-based assays involving the GPCR protein or GPCR protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native GPCR protein is useful for assaying compounds that stimulate or inhibit GPCR protein function.

Host cells are also useful for identifying GPCR protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant GPCR protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native GPCR protein.

Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a GPCR protein and identifying and evaluating modulators of GPCR protein activity. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of

a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the GPCR protein nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the GPCR protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al. PNAS 89*:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science 251*:1351-1355 (1991). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature 385*:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_o phase.

The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, GPCR protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* GPCR protein function, including ligand interaction, the effect of specific mutant GPCR proteins on GPCR protein function and ligand interaction, and the effect of chimeric GPCR proteins. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more GPCR protein functions.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

Claims

That which is claimed is:

1. An isolated peptide consisting of an amino acid sequence selected from the group consisting of:

(a) an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5;

(b) an amino acid sequence of an allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1 (transcript), 4 (transcript), and 3 (genomic);

(c) an amino acid sequence of an ortholog of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1 (transcript), 4 (transcript), and 3 (genomic); and

(d) a fragment of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said fragment comprises at least 10 contiguous amino acids.

2. An isolated peptide comprising an amino acid sequence selected from the group consisting of:

(a) an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5;

(b) an amino acid sequence of an allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1 (transcript), 4 (transcript), and 3 (genomic);

(c) an amino acid sequence of an ortholog of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1 (transcript), 4 (transcript), and 3

(genomic); and

(d) a fragment of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said fragment comprises at least 10 contiguous amino acids.

3. An isolated antibody that selectively binds to a peptide of claim 2.

4. An isolated nucleic acid molecule consisting of a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence that encodes an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5;

(b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1 (transcript), 4 (transcript), and 3 (genomic);

(c) a nucleotide sequence that encodes an ortholog of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1 (transcript), 4 (transcript), and 3 (genomic);

(d) a nucleotide sequence that encodes a fragment of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said fragment comprises at least 10 contiguous amino acids; and

(e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

5. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence that encodes an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5;

(b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1 (transcript), 4 (transcript), and 3 (genomic);

(c) a nucleotide sequence that encodes an ortholog of an amino acid sequence

selected from the group consisting of SEQ ID NOS:2 and 5, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1 (transcript), 4 (transcript), and 3 (genomic);

(d) a nucleotide sequence that encodes a fragment of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said fragment comprises at least 10 contiguous amino acids; and

(e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

6. A gene chip comprising a nucleic acid molecule of claim 5.

7. A transgenic non-human animal comprising a nucleic acid molecule of claim 5.

8. A nucleic acid vector comprising a nucleic acid molecule of claim 5.

9. A host cell containing the vector of claim 8.

10. A method for producing any of the peptides of claim 1 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.

11. A method for producing any of the peptides of claim 2 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.

12. A method for detecting the presence of any of the peptides of claim 2 in a sample, said method comprising contacting said sample with a detection agent that specifically allows detection of the presence of the peptide in the sample and then detecting the presence of the peptide.

13. A method for detecting the presence of a nucleic acid molecule of claim 5 in a sample, said method comprising contacting the sample with an oligonucleotide that hybridizes to

said nucleic acid molecule under stringent conditions and determining whether the oligonucleotide binds to said nucleic acid molecule in the sample.

14. A method for identifying a modulator of a peptide of claim 2, said method comprising contacting said peptide with an agent and determining if said agent has modulated the function or activity of said peptide.

15. The method of claim 14, wherein said agent is administered to a host cell comprising an expression vector that expresses said peptide.

16. A method for identifying an agent that binds to any of the peptides of claim 2, said method comprising contacting the peptide with an agent and assaying the contacted mixture to determine whether a complex is formed with the agent bound to the peptide.

17. A pharmaceutical composition comprising an agent identified by the method of claim 16 and a pharmaceutically acceptable carrier therefor.

18. A method for treating a disease or condition mediated by a human proteases, said method comprising administering to a patient a pharmaceutically effective amount of an agent identified by the method of claim 16.

19. A method for identifying a modulator of the expression of a peptide of claim 2, said method comprising contacting a cell expressing said peptide with an agent, and determining if said agent has modulated the expression of said peptide.

20. An isolated human protease peptide having an amino acid sequence that shares at least 70% homology with an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5.

21. A peptide according to claim 20 that shares at least 90 percent homology with an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5.

22. An isolated nucleic acid molecule encoding a human protease peptide, said nucleic acid molecule sharing at least 80 percent homology with a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1 (transcript), 4 (transcript), and 3 (genomic).

23. A nucleic acid molecule according to claim 22 that shares at least 90 percent homology with a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1 (transcript), 4 (transcript), and 3 (genomic).

FORM 1:

```
       1  ATGACGTCCA CCTGCACCAA CAGCACGCGC GAGAGTAACA GCAGCCACAC
      51  GTGCATGCCC CTCTCCAAAA TGCCCATCAG CCTGGCCCAC GGCATCATCC
     101  GCTCAACCGT GCTGGTTATC TTCCTCGCCG CCTCTTTCGT CGGCAACATA
     151  GTGCTGGCGC TAGTGTTGCA GCGCAAGCCG CAGCTGCTGC AGGTGACCAA
     201  CCGTTTTATC TTTAACCTCC TCGTCACCGA CCTGCTGCAG ATTTCGCTCG
     251  TGGCCCCCTG GGTGGTGGCC ACCTCTGTGC CTCTCTTCTG GCCCCTCAAC
     301  AGCCACTTCT GCACGGCCCT GGTTAGCCTC ACCCACCTGT TCGCCTTCGC
     351  CAGCGTCAAC ACCATTGTCG TGGTGTCAGT GGATCGCTAC TTGTCCATCA
     401  TCCACCCTCT CTCCTACCCG TCCAAGATGA CCCAGCGCCG CGGTTACCTG
     451  CTCCTCTATG GCACCTGGAT TGTGGCCATC CTGCAGAGCA CTCCTCCACT
     501  CTACGGCTGG GGCCAGGCTG CCTTTGATGA GCGCAATGCT CTCTGCTCCA
     551  TGATCTGGGG GGCCAGCCCC AGCTACACTA TTCTCAGCGT GGTGTCCTTC
     601  ATCGTCATTC CACTGATTGT CATGATTGCC TGCTACTCCG TGGTGTTCTG
     651  TGCAGCCCGG AGGCAGCATG CTCTGCTGTA CAATGTCAAG AGACACAGCT
     701  TGGAAGTGCG AGTCAAGGAC TGTGTGGAGA ATGAGGATGA AGAGGGAGCA
     751  GAGAAGAAGG AGGAGTTCCA GGATGAGAGT GAGTTTCGCC GCCAGCATGA
     801  AGGTGAGGTC AAGGCCAAGG AGGGCAGAAT GGAAGCCAAG GACGGCAGCC
     851  TGAAGGCCAA GGAAGGAAGC ACGGGGACCA GTGAGAGTAG TGTAGAGGCC
     901  AGGGGCAGCG AGGAGGTCAG AGAGAGCAGC ACGGTGGCCA GCGACGGCAG
     951  CATGGAGGGT AAGGAAGGCA GCACCAAAGT TGAGGAGAAC AGCATGAAGG
    1001  CAGACAAGGG TCGCACAGAG GTCAACCAGT GCAGCATTGA CTTGGGTGAA
    1051  GATGACATGG AGTTTGGTGA AGACGACATC AATTTCAGTG AGGATGACGT
    1101  CGAGGCAGTG AACATCCCGG AGAGCCTCCC ACCCAGTCGT CGTAACAGCA
    1151  ACAGCAACCC TCCTCTGCCC AGGTGCTACC AGTGCAAAGC TGCTAAAGTG
    1201  ATCTTCATCA TCATTTTCTC CTATGTGCTA TCCCTGGGGC CCTACTGCTT
    1251  TTTAGCAGTC CTGGCCGTGT GGGTGGATGT CGAAACCCAG GTACCCCAGT
    1301  GGGTGATCAC CATAATCATC TGGCTTTTCT TCCTGCAGTG CTGCATCCAC
    1351  CCCTATGTCT ATGGCTACAT GCACAAGACC ATTAAGAAGG AAATCCAGGA
    1401  CATGCTGAAG AAGTTCTTCT GCAAGGAAAA GCCCCCGAAA GAAGATAGCC
    1451  ACCCAGACCT GCCCGGAACA GAGGGTGGGA CTGAAGGCAA GATTGTCCCT
    1501  TCCTACGATT CTGCTACTTT TCCTTGA
```

(SEQ ID NO:1)

FEATURES:
Start: 1
Stop: 1525

FORM 2:

```
   1 ATGACGTCCA CCTGCACCAA CAGCACGCGC GAGAGTAACA GCAGCCACAC
  51 GTGCATGCCC CTCTCCAAAA TGCCCATCAG CCTGGCCCAC GGCATCATCC
 101 GCTCAACCGT GCTGGTTATC TTCCTCGCCG CCTCTTTCGT CGGCAACATA
 151 GTGCTGGCGC TAGTGTTGCA GCGCAAGCCG CAGCTGCTGC AGGTGACCAA
 201 CCGTTTTATC TTTAACCTCC TCGTCACCGA CCTGCTGCAG ATTTCGCTCG
 251 TGGCCCCCTG GGTGGTGGCC ACCTCTGTGC CTCTCTTCTG GCCCCTCAAC
 301 AGCCACTTCT GCACGGCCCT GGTTAGCCTC ACCCACCTGT TCGCCTTCGC
 351 CAGCGTCAAC ACCATTGTCG TGGTGTCAGT GGATCGCTAC TTGTCCATCA
 401 TCCACCCTCT CTCCTACCCG TCCAAGATGA CCCAGCGCCG CGGTTACCTG
 451 CTCCTCTATG GCACCTGGAT TGTGGCCATC CTGCAGAGCA CTCCTCCACT
 501 CTACGGCTGG GGCCAGGCTG CCTTTGATGA GCGCAATGCT CTCTGCTCCA
 551 TGATCTGGGG GGCCAGCCCC AGCTACACTA TTCTCAGCGT GGTGTCCTTC
 601 ATCGTCATTC CACTGATTGT CATGATTGCC TGCTACTCCG TGGTGTTCTG
 651 TGCAGCCCGG AGGCAGCATG CTCTGCTGTA CAATGTCAAG AGACACAGCT
 701 TGGAAGTGCG AGTCAAGGAC TGTGTGGAGA ATGAGGATGA AGAGGGAGCA
 751 GAGAAGAAGG AGGAGTTCCA GGATGAGATG AACATCCCGG AGAGCCTCCC
 801 ACCCAGTCGT CGTAACAGCA ACAGCAACCC TCCTCTGCCC AGGTGCTACC
 851 AGTGCAAAGC TGCTAAAGTG ATCTTCATCA TCATTTTCTC CTATGTGCTA
 901 TCCCTGGGGC CCTACTGCTT TTTAGCAGTC CTGGCCGTGT GGGTGGATGT
 951 CGAAACCCAG GTACCCCAGT GGGTGATCAC CATAATCATC TGGCTTTTCT
1001 TCCTGCAGTG CTGCATCCAC CCCTATGTCT ATGGCTACAT GCACAAGACC
1051 ATTAAGAAGG AAATCCAGGA CATGCTGAAG AAGTTCTTCT GCAAGGAAAA
1101 GCCCCCGAAA GAAGATAGCC ACCCAGACCT GCCCGGAACA GAGGGTGGGA
1151 CTGAAGGCAA GATTGTCCCT TCCTACGATT CTGCTACTTT TCCTTGA
```

(SEQ ID NO:4)

FEATURES:
Start: 1
Stop: 1195

FIGURE 1

HOMOLOGOUS PROTEIN:
Top 10 BLAST Hits:

	Score	E
CRA|18000005170915|gi|6677701|ref|NP_031395.1|| G-protein coupl...	130	5e-29
CRA|18000005170915|gi|6677701|gb|AAC61598.1| (AF091890) G-prote...	130	5e-29
CRA|18000005086340|gi|3023219|sp|O02824|A1AA_RABIT ALPHA-1A ADR...	115	1e-24
CRA|18000005086340|gi|3023219|gb|AAB61334.1| (U81982) alpha 1a-...	115	1e-24
CRA|18000004999268|gi|2119508|pir||S65656 alpha-1C-adrenergic r...	113	4e-24
CRA|18000004903730|gi|4501961|ref|NP_000671.1|| adrenergic, alp...	113	4e-24
CRA|18000004976627|gi|666893|gb|AAB59486.1| (L31774) alpha-1C-a...	113	4e-24
CRA|18000004999267|gi|2119507|pir||S65657 alpha-1C-adrenergic r...	113	4e-24
CRA|18000004922459|gi|1168246|sp|P35348|A1AA_HUMAN ALPHA-1A ADR...	113	4e-24
CRA|18000004923366|gi|547222|gb|AAB31165.1| alpha adrenergic re...	113	4e-24

EXPRESSION INFORMATION FOR MODULATORY USE:
Human Brain

FIGURE 1A

```
FORM 1:
       1  MTSTCTNSTR ESNSSHTCMP LSKMPISLAH GIIRSTVLVI FLAASFVGNI
      51  VLALVLQRKP QLLQVTNRFI FNLLVTDLLQ ISLVAPWVVA TSVPLFWPLN
     101  SHFCTALVSL THLFAFASVN TIVVVSVDRY LSIIHPLSYP SKMTQRRGYL
     151  LLYGTWIVAI LQSTPPLYGW GQAAFDERNA LCSMIWGASP SYTILSVVSF
     201  IVIPLIVMIA CYSVVFCAAR RQHALLYNVK RHSLEVRVKD CVENEDEEGA
     251  EKKEEFQDES EFRRQHEGEV KAKEGRMEAK DGSLKAKEGS TGTSESSVEA
     301  RGSEEVRESS TVASDGSMEG KEGSTKVEEN SMKADKGRTE VNQCSIDLGE
     351  DDMEFGEDDI NFSEDDVEAV NIPESLPPSR RNSNSNPPLP RCYQCKAAKV
     401  IFIIIFSYVL SLGPYCFLAV LAVWVDVETQ VPQWVITIII WLFFLQCCIH
     451  PYVYGYMHKT IKKEIQDMLK KFFCKEKPPK EDSHPDLPGT EGGTEGKIVP
     501  SYDSATFP
             (SEQ ID NO:2)
FORM 2:
    1 MTSTCTNSTR ESNSSHTCMP LSKMPISLAH GIIRSTVLVI FLAASFVGNI
   51 VLALVLQRKP QLLQVTNRFI FNLLVTDLLQ ISLVAPWVVA TSVPLFWPLN
  101 SHFCTALVSL THLFAFASVN TIVVVSVDRY LSIIHPLSYP SKMTQRRGYL
  151 LLYGTWIVAI LQSTPPLYGW GQAAFDERNA LCSMIWGASP SYTILSVVSF
  201 IVIPLIVMIA CYSVVFCAAR RQHALLYNVK RHSLEVRVKD CVENEDEEGA
  251 EKKEEFQDEM NIPESLPPSR RNSNSNPPLP RCYQCKAAKV IFIIIFSYVL
  301 SLGPYCFLAV LAVWVDVETQ VPQWVITIII WLFFLQCCIH PYVYGYMHKT
  351 IKKEIQDMLK KFFCKEKPPK EDSHPDLPGT EGGTEGKIVP SYDSATFP
             (SEQ ID NO:5)
FEATURES:
Functional domains and key regions:
[1] PDOC00001 PS00001 ASN_GLYCOSYLATIONN-glycosylation site
Number of matches: 3
      1       7-10 NSTR
      2      13-16 NSSH
      3    361-364 NFSE
-------------------------------------------------------------------------------
[2] PDOC00004 PS00004 CAMP_PHOSPHO_SITEcAMP- and cGMP-dependent protein kinase
phosphorylation site
Number of matches: 2
      1    230-233 KRHS
      2    380-383 RRNS
-------------------------------------------------------------------------------
[3] PDOC00005 PS00005 PKC_PHOSPHO_SITEProtein kinase C phosphorylation site
Number of matches: 8
      1       8-10 STR
      2      66-68 TNR
      3    144-146 TQR
      4    283-285 SLK
      5    324-326 STK
      6    331-333 SMK
      7    379-381 SRR
      8    460-462 TIK
-------------------------------------------------------------------------------
[4] PDOC00006 PS00006 CK2_PHOSPHO_SITECasein kinase II phosphorylation site
Number of matches: 5
      1       8-11 STRE
      2    296-299 SSVE
      3    325-328 TKVE
      4    363-366 SEDD
      5    483-486 SHPD
-------------------------------------------------------------------------------
[5] PDOC00008 PS00008 MYRISTYLN-myristoylation site
Number of matches: 8
      1      31-36 GIIRST
      2    187-192 GASPSY
      3    282-287 GSLKAK
      4    289-294 GSTGTS
      5    292-297 GTSESS
      6    316-321 GSMEGK
      7    489-494 GTEGGT
      8    492-497 GGTEGK
-------------------------------------------------------------------------------
[6] PDOC00210 PS00237 G_PROTEIN_RECEPTORG-protein coupled receptors signature
          117-133 ASVNTIVVVSVDRYLSI
```

FIGURE 2, page 1 of 4

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	36	56	1.964	Certain
2	79	99	1.652	Certain
3	102	122	1.564	Certain
4	148	168	1.329	Certain
5	193	213	2.334	Certain
6	400	420	1.967	Certain
7	431	451	1.854	Certain

BLAST Alignment to Top Hit:

>CRA|18000005170915|gi|6677701|ref|NP_031395.1|| G-protein coupled receptor
Length = 407

Score = 130 bits (323), Expect = 5e-29
Identities = 63/192 (32%), Positives = 107/192 (54%), Gaps = 2/192 (1%)

```
Query: 8   GIIRSTVLVIFLAASFV--GNIVLALVLQRKPQLLQVTNRFIFNLLVTDLLQISLVAPWV 65
           G+I +  + I +  . FV  GN+V+ + L +K  LL ++N+F+F+L +++ L  LV P+V
Sbjct: 25  GVIITQFIAIIVITIFVCLGNLVIVVTLYKKSYLLTLSNKFVFSLTLSNFLLSVLVLPFV 84

Query: 66  VATSVPLFWPLNSHFCTALVSLTHLFAFASVNTIVVVSVDRYLSIIHPLSYPSKMTQRRG 125
           V +S+   W     +C     L  L + AS+ T+ V+++DRY ++++P+ YP K+T  R
Sbjct: 85  VTSSIRREWIFGVVWCNFSALLYLLISSASMLTLGVIAIDRYYAVLYPMVYPMKITGNRA 144

Query: 126 YLLLYGTWIVAILQSTPPLYGWGQAAFDERNALCSMIWGASPSYTILSVVSFIVIPLIVM 185
            + L   W+ +++   PPL+GW    FDE   +C   W   P YT    +   + P +VM
Sbjct: 145 VMALVYIWLHSLIGCLPPLFGWSSVEFDEFKWMCVAAWHREPGYTAFWQIWCALFPFLVM 204

Query: 186 IACYSVVFCAAR 197
           + CY  +F  AR
Sbjct: 205 LVCYGFIFRVAR 216 (SEQ ID NO :6)
```

Score = 48.4 bits (113), Expect = 2e-04
Identities = 25/76 (32%), Positives = 43/76 (55%), Gaps = 3/76 (3%)

```
Query: 371 QCKAAKVIFIIIFSYVLSLGPYCF-LAVLAVWVDVETQVPQWVITIIIWLFFLQCCIHPY 429
           QCKA   I +++ +++++ GPY   +A  A+W   ++ V   + T   WL F     HP
Sbjct: 265 QCKALITILVVLGAFMVTWGPYMVVIASEALWG--KSSVSPSLETWATWLSFASAVCHPL 322

Query: 430 VYGYMHKTIKKEIQDM 445
           +YG  +KT++KE+  M
Sbjct: 323 IYGLWNKTVRKELLGM 338  (Seq ID NO:7)
```

ALIGNMENT OF FORM 1 AND FORM 2:

>FORM 1
Length = 508

Score = 521 bits (1327), Expect = e-152
Identities = 259/259 (100%), Positives = 259/259 (100%)

```
FORM 2: 1   MTSTCTNSTRESNSSHTCMPLSKMPISLAHGIIRSTVLVIFLAASFVGNIVLALVLQRKP 60
           MTSTCTNSTRESNSSHTCMPLSKMPISLAHGIIRSTVLVIFLAASFVGNIVLALVLQRKP
FORM 1: 1   MTSTCTNSTRESNSSHTCMPLSKMPISLAHGIIRSTVLVIFLAASFVGNIVLALVLQRKP 60

FORM 2: 61  QLLQVTNRFIFNLLVTDLLQISLVAPWVVATSVPLFWPLNSHFCTALVSLTHLFAFASVN 120
           QLLQVTNRFIFNLLVTDLLQISLVAPWVVATSVPLFWPLNSHFCTALVSLTHLFAFASVN
FORM 1: 61  QLLQVTNRFIFNLLVTDLLQISLVAPWVVATSVPLFWPLNSHFCTALVSLTHLFAFASVN 120

FORM 2: 121 TIVVVSVDRYLSIIHPLSYPSKMTQRRGYLLLYGTWIVAILQSTPPLYGWGQAAFDERNA 180
           TIVVVSVDRYLSIIHPLSYPSKMTQRRGYLLLYGTWIVAILQSTPPLYGWGQAAFDERNA
FORM 1: 121 TIVVVSVDRYLSIIHPLSYPSKMTQRRGYLLLYGTWIVAILQSTPPLYGWGQAAFDERNA 180

FORM 2: 181 LCSMIWGASPSYTILSVVSFIVIPLIVMIACYSVVFCAARRQHALLYNVKRHSLEVRVKD 240
           LCSMIWGASPSYTILSVVSFIVIPLIVMIACYSVVFCAARRQHALLYNVKRHSLEVRVKD
```

```
FORM 1: 181 LCSMIWGASPSYTILSVVSFIVIPLIVMIACYSVVFCAARRQHALLYNVKRHSLEVRVKD 240

FORM 2: 241 CVENEDEEGAEKKEEFQDE 259
            CVENEDEEGAEKKEEFQDE
FORM 1: 241 CVENEDEEGAEKKEEFQDE 259


 Score =  303 bits (767), Expect = 1e-86
 Identities = 140/152 (92%), Positives = 144/152 (94%)

FORM 2: 247 EEGAEKKEEFQDEMNIPESLPPSRRNSNSNPPLPRCYQCKAAKVIFIIIFSYVLSLGPYC 306
            E+     E+  + +NIPESLPPSRRNSNSNPPLPRCYQCKAAKVIFIIIFSYVLSLGPYC
FORM 1: 357 EDDINFSEDDVEAVNIPESLPPSRRNSNSNPPLPRCYQCKAAKVIFIIIFSYVLSLGPYC 416

FORM 2: 307 FLAVLAVWVDVETQVPQWVITIIIWLFFLQCCIHPYVYGYMHKTIKKEIQDMLKKFFCKE 366
            FLAVLAVWVDVETQVPQWVITIIIWLFFLQCCIHPYVYGYMHKTIKKEIQDMLKKFFCKE
FORM 1: 417 FLAVLAVWVDVETQVPQWVITIIIWLFFLQCCIHPYVYGYMHKTIKKEIQDMLKKFFCKE 476

FORM 2: 367 KPPKEDSHPDLPGTEGGTEGKIVPSYDSATFP 398
            KPPKEDSHPDLPGTEGGTEGKIVPSYDSATFP
FORM 1: 477 KPPKEDSHPDLPGTEGGTEGKIVPSYDSATFP 508
```

Hmmer search results (Pfam):

FORM 1:

Model	Description	Score	E-value	N
PF00001	7 transmembrane receptor (rhodopsin family)	196.1	4.7e-61	1
CE00503	CE00503 ADRENERGIC_RECEPTOR	33.2	1.7e-08	1
CE00326	E00326 mu_opioid_receptor	30.4	3.1e-08	1
CE00267	E00267 Octopamine_receptor	25.2	2.5e-06	2
CE00266	CE00266 Conopressin_receptor	15.4	0.0023	1
CE00179	CE00179 Burkitts_lymphoma_receptor	13.9	0.0048	1
CE00323	E00323 kappa_opioid_receptor	13.4	0.0021	1
CE00513	CE00513 NEUROPEPTIDE_Y_RECEPTOR	13.2	0.004	2
CE00328	E00328 tachykinin_receptor	13.0	0.0066	2
CE00340	E00340 thrombin_receptor	12.6	0.0082	1

FORM 2:

Model	Description	Score	E-value	N
PF00001	7 transmembrane receptor (rhodop	177.9	2.5e-55	1
CE00503	CE00503 ADRENERGIC_RECEPTOR	33.2	1.7e-08	1
CE00326	E00326 mu_opioid_receptor	30.4	3.1e-08	1
CE00267	E00267 Octopamine_receptor	25.2	2.5e-06	2
CE00266	CE00266 Conopressin_receptor	15.4	0.0023	1
CE00179	CE00179 Burkitts_lymphoma_recept	13.9	0.0048	1
CE00323	E00323 kappa_opioid_receptor	13.4	0.0021	1
CE00513	CE00513 NEUROPEPTIDE_Y_RECEPTOR	13.2	0.004	2
CE00328	E00328 tachykinin_receptor	13.0	0.0066	2
CE00340	E00340 thrombin_receptor	12.6	0.0082	1
CE00531	CE00531 CHEMOKINE_RECEPTOR_TYPE_	12.1	0.012	2
CE00336	E00336 orexin_receptor	11.6	0.0094	1
CE00175	CE00175 Purinoceptors	9.2	0.11	2
CE00507	CE00507 DOPAMINE_RECEPTOR	8.6	0.11	1
CE00516	CE00516 SEROTONIN_RECEPTOR	7.8	0.18	1
CE00511	CE00511 MELANOCORTIN_RECEPTOR	7.3	0.3	1
CE00529	CE00529 CHEMOKINE_RECEPTOR_TYPE_	7.1	0.15	1
CE00177	CE00177 ADENOSINE	7.1	0.39	3
CE00505	CE00505 CHEMOKINE_RECEPTOR	6.4	0.56	1
CE00205	CE00205 GH_secretagogue	4.8	0.2	1
CE00343	E00343 prostaglandin_receptor_ep	4.6	1.1	1
CE00519	CE00519 BRADYKININ_RECEPTOR	4.2	2.7	1
CE00337	E00337 gonadotropin_releasing_ho	3.8	4.8	2
CE00517	CE00517 SOMATOSTATIN_RECEPTOR	3.6	4	1
CE00139	E00139 Cannabis	3.6	2.1	1
CE00530	CE00530 CHEMOKINE_RECEPTOR_TYPE_	2.5	3.4	1
CE00506	CE00506 CHOLECYSTOKININ_RECEPTOR	1.7	6	1

CE00321	E00321 galanin_receptor	1.4	9.8	1
PF01821	Anaphylotoxin-like domain	0.9	9.5	1
CE00521	CE00521 THROMBIN_RECEPTOR	-0.0	7.9	1
CE00318	E00318 MIP-1_receptor	-78.6	0.58	1
olfactory_receptor_2	CE00540 olfactory_receptor_2	-237.6	0.029	1
olfactory_receptor_5	CE00542 olfactory_receptor_5	-285.2	1.4	1

Parsed for domains:

Model	Domain	seq-f	seq-t		hmm-f	hmm-t		score	E-value
CE00506	1/1	39	79	..	63	102	..	1.7	6
CE00513	1/2	47	79	..	106	138	..	0.1	22
CE00529	1/1	48	135	..	70	156	..	7.1	0.15
CE00179	1/1	39	135	..	62	155	..	13.9	0.0048
CE00328	1/2	116	137	..	188	209	..	17.0	0.0005
CE00321	1/1	110	137	..	126	154	..	1.4	9.8
CE00521	1/1	116	137	..	191	212	..	-0.0	7.9
CE00517	1/1	124	138	..	141	155	..	3.6	4
CE00323	1/1	47	139	..	85	179	..	13.4	0.0021
CE00343	1/1	73	154	..	78	168	..	4.6	1.1
CE00326	1/1	29	157	..	1	132	[.	30.4	3.1e-08
CE00205	1/1	116	159	..	128	171	..	4.8	0.2
CE00336	1/1	38	160	..	54	173	..	11.6	0.0094
CE00340	1/1	116	161	..	191	236	..	12.6	0.0082
CE00511	1/1	127	164	..	141	178	..	7.3	0.3
CE00531	1/2	46	165	..	47	164	..	18.8	0.00015
CE00513	2/2	118	165	..	178	223	..	13.3	0.0035
CE00266	1/1	117	165	..	83	131	..	15.4	0.0023
CE00519	1/1	116	167	..	122	173	..	4.2	2.7
CE00175	1/2	126	168	..	106	148	..	11.8	0.02
CE00505	1/1	116	168	..	119	171	..	6.4	0.56
CE00337	1/2	123	168	..	150	197	..	5.9	1.3
CE00507	1/1	124	170	..	142	192	..	8.6	0.11
CE00267	1/2	125	170	..	126	171	..	27.9	4.5e-07
CE00177	1/3	100	170	..	77	147	..	8.2	0.18
CE00139	1/1	127	170	..	129	172	..	3.6	2.1
CE00503	1/1	124	173	..	148	198	..	33.2	1.7e-08
CE00531	2/2	197	213	..	201	217	..	-3.0	2.1e+02
CE00175	2/2	203	215	..	188	200	..	-4.0	6.6e+02
CE00530	1/1	116	215	..	122	225	..	2.5	3.4
CE00328	2/2	203	215	..	288	300	..	0.2	22
CE00516	1/1	35	222	..	148	344	..	7.8	0.18
CE00337	2/2	192	222	..	233	265	..	0.1	51
CE00267	2/2	202	222	..	210	230	..	2.2	8.3
CE00177	2/3	200	222	..	195	217	..	4.2	2.4
olfactory_receptor_2	1/1	24	241	..	1	313	[]	-237.6	0.029
olfactory_receptor_5	1/1	85	245	..	1	308	[]	-285.2	1.4
CE00318	1/1	13	336	..	1	346	[]	-78.6	0.58
PF01821	1/1	333	338	..	32	37	.]	0.9	9.5
PF00001	1/1	48	344	..	1	259	[]	177.9	2.5e-55
CE00177	3/3	286	374	..	237	327	.]	5.2	1.3

FIGURE 2, page 4 of 4

```
   1 ATCTCACAAG GCTGGGGTAA AGATTCAGCG GTATTCTCTG TGTAAACAGT
  51 TTAACACAGT ACCTGGCAGA GAGAGGGGGA GAAAGAGGGA GAGAGGGAGA
 101 GATGCTGACT GACTTACAGA CAGACATACT TCAATCAAAG AATATAGTGG
 151 GTAATTATGT GGCTTGTCTT AAATGAAATG CTTCCAAATC CCACAGGGAA
 201 AGCACTGCCG CTTAGCTCTT CCTGGGAGCT AGAGAAATTC GTGGAAAATC
 251 GTTGACATTA AAGTCACGAT TCCTTTCTCG GGCTCAGGGG CTATTCATAC
 301 TTTGGAGGGT TTTCAGAAGT CGAGGGGAGT CCCGGGCCCT CCTATTTGGA
 351 AGAATCGGAA GGATAGTGCA GGGTCTGTGC CTGGGAGCTG GCCCGAAATC
 401 AGGGGGCTAG GAAGGCACGG GGTTTGCACG AAAATCAGCC CCGGTGACAA
 451 TCCCCAGTCT GTGAGCTCTC GGTGGACTCG GGGCCCCAGG GGTGGGAACC
 501 CGGAGTGAGA AAATCTGAGT GGGGTCTGGT GCACGGCCCC CCACCTGCCA
 551 GCCCCGCTCC CTGGGCTGCG GCGGCCTCGT GCGTCACGGG CGCAGGCGGG
 601 GGCCGAGGGG AGGGCGGCAC CGCAAGGTGA GGGCACCGCG CTGCCTTAAG
 651 ACGCCGCGCG CCCCCGCTGC CTCTCAGAGC TTGCAGCGCG GCGCGCTGGC
 701 CGGACGTGCG GGCACAGACG CTGTGCCCCG CTCCTGCCGC GCCCGCAGCG
 751 CCCGCTGCCC CGTGCCAGCT CGCAGCCGCC CGAGCCCCGA GAGGCGGCCG
 801 GGACGAGCGC GAGGAGCCCC GAAGGACCTG CACGCGGCGC CCGCGACGCC
 851 ACGTCTGCAG CAGCAACAGG TAAGCAGCCC GGCCCTACTC CGGGCTCCGT
 901 GGCCCCCAGC ACCCGGCAGG GCCCCGAGCC TGGGAGCCCA CGACGCCCCC
 951 TCTGAAGCGT CCCGACGGCT GCTGAGCTAC CAGGCGCGGT GGCCCGCGGT
1001 GCGGCCCCTT CTCGCCTTTC CCTAGCCATC CTCCCGCCCT CGCGTTTTCT
1051 CTCTCACTTT TTCTCGTGCG TTCTTTTCCC CCTTCCTCTC CATTCTCTCG
1101 TTTTTTCCTC ATCCTTTCCC TCTTCCTCTG CTTGTTCTCT TTTCTTATTA
1151 TTTTTTGCTT TTTCTGTCTG CGTCTCTCCC AGTCTGTCTT GCTCTCTCAT
1201 TCTCTCTCAC TCTCACGCTC CCATTTCTTC TTTTTTCTTT CTTTCTCCCT
1251 CTCTGTTGCT CACATTCTCT CTGTCTCTCT GTATGTATCT CATTCTTTCG
1301 CTGTCTCTCA TTTCCTTTTT GTCTCTCTGC CTCTCAGTCT CTCTCGCCCT
1351 GCCTTTCTGA CTCTCTCCAC CACCCTGCAC CCCTGCCTCT TTTGCCTTTC
1401 TCCCCAGCTC CTTTCCCACT CCCCTCGGCT GTCACCGACA CCAAGCAGAT
1451 CCGGAGAAAG CAAAGAGGAT GCCAAATTGG GAATGGGGAT GGGGCTGTGG
1501 AGGGTGAGGG TAGGGGAGCT CAATAGGCAT CCTCTCTGCA TTTTCCAGTC
1551 TCCATTGCAC TGCTGATAAA CTAAGGGTTT TTTAACCAAA GCCCGCGGTG
1601 GCCGGGGCGG CAGGTGGTTT CCACGGCAAA ATAAAGTCCA CAGGGAGACA
1651 GAGATCCAAA GAGACAGAGA TCCAGCCAGA GAAAAACCAA GGTGAGGGGC
1701 CATGGGAAAA AGATGTAGAG AAATAGAGAT GCGCACACAC TCAAAGATAC
1751 AAAGCAAGTG TTTAGCCAGA GTCCTGGGGT GACAGGAAAG TGAAGGACAG
1801 AGAAACCAAC GAGAGAGTGC ACACACTAAA GAGATAGCTT GAGAGACAGG
1851 TGAGACAAGC AGTAGAGTGT TAGAGATCAA TGGAGAGAGA GGCAGAGTAA
1901 TGCAGAGAGA ACTTAGGCAG GGAGAGACAG AAACCACACG GAGAGACACA
1951 GAGACAAGAG GGTCCGAGAA AGACAGAGCA AGCGCACACG TGCCAGTGAG
2001 AGACAGAAGG AGACACCCAG AGAACCCGAG GGAGACGCAG AAACAAGAGA
2051 GTTGGAGAGA CACGGAGACG GGCAGAGACA CAGCGAACGA CATTGAACAG
2101 GTGCGCTAGG GACCCACAAA GATGAACTGC TGGGGAAATA GACTGAGAGA
2151 GAAATGGAGA GACAAGAGAG CACTAGGGAA AGACGGAGAC ACACATGAAC
2201 TAGCGAGCGA CACAGTGACA GGCAGAGACA CAGAGAGACA GAATGGATCA
2251 GAAAGAGACT GAGGGAGAGA CGGGCAAGAC TGTGAGAAAG ATAGGGAGCG
2301 AGAGCGCAAG CCCTGCAGAG AAAGTGCGCT AGCGAAAGAG ACACACACAG
2351 AGAGACTGAG TGACAAACAG AGTGGAAGAG ATAAATGAGA GTAACAGGAA
2401 AGAGACAGCG TGCACGTGCT ATCGAGAGAC GGAGAAACAG AGACAGACAT
2451 AAGTGAGAGA CAAGGATGAG AAAGAATCAG GAAACGGTGA GACACATACA
2501 GACACACAAC CCGAGAGATA AAGGAAAAAA AGTGCGCTAG CGAGAGACAC
2551 TGGGAGAAGA GACAGGAAAC TGACGGTCGT AAACAGACAG AAACAGAGAG
2601 CGCACGCGCA TGCGCTAGCG CTAGCGACCA AAACTCCCCG GAGCCAGAGA
2651 CAGTGTGAGA GACAAGCAGA GAAAGCGCAC GCACGCACGC GCCAGCAGGA
2701 GAAACAGATG AGAGGAAATC AGAGCCCTGG AGAGAGACAG GCAGACAGAT
2751 CTGGAGAGTC CGGAAAGGAG CCATAGAAGC TGCCCGCACT GGGGATGGAG
2801 CCGTGCGGAA ACCCGGGGTA GGGGGTCCTG CAGCGTCCTT GCTGGGCGCG
2851 GAGGCTTCTC CCCTTGACGG GTGACTAACT CTGCCTGCGT GTTTCTTTTG
2901 TCACCAGCAT AGGCACTGAG TGCGGTCTGT GCACCCCTTT GCCACCCACC
2951 GGTGCCGGCA CTGAGCCTGC AACCTGTCTC ACGCCCTCTG GCTGTTGCCA
3001 TGACGTCCAC CTGCACCAAC AGCACGCGCG AGAGTAACAG CAGCCACACG
3051 TGCATGCCCC TCTCCAAAAT GCCCATCAGC CTGGCCCACG GCATCATCCG
3101 CTCAACCGTG CTGGTTATCT TCCTCGCCGC CTCTTTCGTC GGCAACATAG
3151 TGCTGGCGCT AGTGTTGCAG CGCAAGCCGC AGCTGCTGCA GGTGACCAAC
3201 CGTTTTATCT TTAACCTCCT CGTCACCGAC CTGCTGCAGA TTTCGCTCGT
3251 GGCCCCCTGG GTGGTGGCCA CCTCTGTGCC TCTCTTCTGG CCCCTCAACA
```

FIGURE 3, page 1 of 4

```
3301 GCCACTTCTG CACGGCCCTG GTTAGCCTCA CCCACCTGTT CGCCTTCGCC
3351 AGCGTCAACA CCATTGTCGT GGTGTCAGTG GATCGCTACT TGTCCATCAT
3401 CCACCCTCTC TCCTACCCGT CCAAGATGAC CCAGCGCCGC GGTTACCTGC
3451 TCCTCTATGG CACCTGGATT GTGGCCATCC TGCAGAGCAC TCCTCCACTC
3501 TACGGCTGGG GCCAGGCTGC CTTTGATGAG CGCAATGCTC TCTGCTCCAT
3551 GATCTGGGGG GCCAGCCCCA GCTACACTAT TCTCAGCGTG GTGTCCTTCA
3601 TCGTCATTCC ACTGATTGTC ATGATTGCCT GCTACTCCGT GGTGTTCTGT
3651 GCAGCCCGGA GGCAGCATGC TCTGCTGTAC AATGTCAAGA GACACAGCTT
3701 GGAAGTGCGA GTCAAGGACT GTGTGGAGAA TGAGGATGAA GAGGGAGCAG
3751 AGAAGAAGGA GGAGTTCCAG GATGAGAGTG AGTTTCGCCG CCAGCATGAA
3801 GGTGAGGTCA AGGCCAAGGA GGGCAGAATG GAAGCCAAGG ACGGCAGCCT
3851 GAAGGCCAAG GAAGGAAGCA CGGGGACCAG TGAGAGTAGT GTAGAGGCCA
3901 GGGGCAGCGA GGAGGTCAGA GAGAGCAGCA CGGTGGCCAG CGACGGCAGC
3951 ATGGAGGGTA AGGAAGGCAG CACCAAAGTT GAGGAGAACA GCATGAAGGC
4001 AGACAAGGGT CGCACAGAGG TCAACCAGTG CAGCATTGAC TTGGGTGAAG
4051 ATGACATGGA GTTTGGTGAA GACGACATCA ATTTCAGTGA GGATGACGTC
4101 GAGGCAGTGA ACATCCCGGA GAGCCTCCCA CCCAGTCGTC GTAACAGCAA
4151 CAGCAACCCT CCTCTGCCCA GGTGCTACCA GTGCAAAGCT GCTAAAGTGA
4201 TCTTCATCAT CATTTTCTCC TATGTGCTAT CCCTGGGGCC CTACTGCTTT
4251 TTAGCAGTCC TGGCCGTGTG GGTGGATGTC GAAACCCAGG TACCCCAGTG
4301 GGTGATCACC ATAATCATCT GGCTTTTCTT CCTGCAGTGC TGCATCCACC
4351 CCTATGTCTA TGGCTACATG CACAAGACCA TTAAGAAGGA AATCCAGGAC
4401 ATGCTGAAGA AGTTCTTCTG CAAGGAAAAG CCCCCGAAAG AAGATAGCCA
4451 CCCAGACCTG CCCGGAACAG AGGGTGGGAC TGAAGGCAAG ATTGTCCCTT
4501 CCTACGATTC TGCTACTTTT CCTTGAAGTT AGTTCTAAGG CAAACCTTGA
4551 ACTGTCCATA ACACGAGAAA CAAGAGGAGA TTTCTTTTCA ATGGACCCAC
4601 AATTCATTAA TGCCAAACCA TACCATTTCA GGCAAAGGTG TTGCACACAC
4651 ATGCTCTTCA CCACAAGGTA GATAAATATA TAGAAGAGGC AGGAACTGGG
4701 GTCTTTCCGT AAAAGCATGG ACTTGAGGAT TCTGACTGAA ATTTTCCCCC
4751 CAAAGATTAT TAGGCTCTAC ATTTCTTAAA GCAACAAGGG CTATCCATTT
4801 TGGACTTGTA GTTGGTATTC TATCTTTTCC AGAGCTACAA CATGCCAACT
4851 TTAGCTCTGA AGGAAAGGGA AGATGATGCT TGTGAACTTA AGGACTTTTC
4901 GGCCCTCGGG TCGGGAGCTC ATGGGCCAGA GCTACAGCTT GTGTTCAACT
4951 GAAAGAAAGG CAATGGACCA AATCATTCAT GGAGCCCAAG AAACAGAACC
5001 TAATGGACTG ATCAACATAT GAGCCAAATT CTGAACTGAA CAGCCCCACA
5051 GTCGGGTGCA AAGACTGTTA CACAAACTAA AACAAAGGGC CTCCTACAGT
5101 TAGAATCTCA AGAAGGTTTC TAGATCCCCT AAAGGGATCC AGAAAAGTAG
5151 AAGGACATGT ATGAAATGGG AAGCTAGTCC AAGGGAAAAG ATTGAGAAAT
5201 AACACACATC TGGAGAGCTA AACAGTTGAC TTTTTTTCCT ATAAAATCTT
5251 GGGTTTATGC ATGGGCTGGA ACTGAGGTCA TTAAGTGTAA ATTGTCAATT
5301 GACACAAATA TTTTCTGTCT CCTGTTTGAA TAATAGTGGG GCAGAAATCA
5351 TGCCACTATT TTACAACTTC CCTTATGTGA CTGAATTGAG ATGCTGGTGG
5401 GAATTCTTCA GATCTCTGCC AACACTTCTG TTTTCTTTTG GTTTGTTTTT
5451 GTCAAATAAG CCTTTTTTAG TCAAACAGTA TTTACAGAAA AAAAGAAATT
5501 CAACTAGAAG TGGCCTAAGT CCTACAAAAT TCATGATGTC ACTGAGGAAT
5551 AATTTGTTCA TCAGAAATAT ATTTTGTGTC CATGAGATCA TAGACAATAA
5601 ATGTGATCTC CACATGGGGA GCAAGGAAGG CAGAATGAAC ATTTTTCTTC
5651 CTCCAGGCAC ACCCATGTGT CTTTTCCACC TGTGGCTCTC TTTAAAGCTT
5701 TTAAGCTCTC TGCAGATGTG AAAGAGAAAT ATCAGAGAGT CAGAAATGAC
5751 AAAGAGGATG ATTTCACAAT ACCTAGAAAA CATGTAACCT ATTCCAAACA
5801 GTCCTAAAAT CAGAGCATTC AGATCAGACA TATCCTAATT AATGCTGTTG
5851 AAATAAATCA CGTTGGGAAA ACTTTAACAA TATCTAAATT ATCCCTAGGG
5901 TCAATTCACA GGAACATTCC TCAAATCCCA AACCGCAAAA TAACTTTGGG
5951 CAGGGATATA CATATACATT TCTGAGGGCA TGGGCCGTAT GTATGTGACC
6001 AAGTAACATG GAACCAAAAA AACAGTCAAG CCAGTGTTTT TGATCCTCCT
6051 ACAGAACAAG TTAAAGCAAC TCCAGAGTCA ACCAACTGTT CATGCAGAAA
6101 TCCACTGTCA ATATGGGTGG AGGGAGTGTT TGGTTGAAAA TGGTTAATCA
6151 GGTAGTTGTA TGATGTAAGA TGACCATCTT CAGAGTTTAG CCTCATTCTT
6201 GTGTGATTGT CATGCCTTTA TTAGAACTCA AGTTTCATTT AAATAAATGC
6251 CCAGCTCATT TATTTTTTTT TATCTCTTCC TCCTCACAGA TTTCAACATG
6301 AACTTCTCAA GGGGTAAACA GCAATGTATT TGGACTGTGA ATAACTCTGC
6351 ATGGGAATTT GGGATTGCCA TGTTCAGAAT TTAGGAAAGT AGAGGGAATA
6401 GAACCAAATA ATATAGAGCT GACCCATCCA ATAAAAATAC CATGATAACC
6451 TTATTAATTC CAACTCCATT AATTTGGAAC TTGTAGTTAT TCAGACAAGG
6501 CATGGGGGCT AAAGTTTACC CTTACACTAT CATTTATTTT TCTACCAAAA
6551 TGCATAAAGT GAATTAACAG TCATAAATTT GTTCTACCAA TTTATTGCCA
```

FIGURE 3, page 2 of 4

```
 6601 AACTCTTGAA CTGACCTTTC CTTAAAGGAT ATCTGGGTGA AGAAATGGCC
 6651 TATGTGATCA ATCCTCCTAC AGGGGAGGGG CAGTGCCTTC AGGTCTGTTC
 6701 AAATTGTCAA AGGAGTTCAA GGTAGCTATA GCCTATCCTT TGAGTAGAAA
 6751 TGCTTACTTG GGTAGGAAAC AGGATTTCAA ACATAAATGT CTCCAAACAT
 6801 TGTGTTAAAA CTGAACTTCT TGTTTTATTT TTAAAGCTCA CCCCCTTCAA
 6851 AGTGTATCAG AGAAAATTGT TTGCCAAAAT CTCAAATCAA AATGGAACCA
 6901 GAACCTGTAC AAGTATGGTA AGTCCATTTA ACAGCACACA CAACATTCTC
 6951 AAGGGCACTG AGATTCCCTT TCCTTTTTGA AGTTCCTCTT TTCCCTATTT
 7001 TAGTCATTGT CCATTATTTT GGTAAATTGG ATTCCTCAAA AGTGAAGACC
 7051 TTTTGAAACT ATGAGCCTGG AAAAGAGGAC CTTTTAATTA AGAGCATTCT
 7101 GCTTTGATGA CATTTTCTTC TAAATGAACA ATAAGGCCTA TGGTTAGCTT
 7151 GGAGATAGCA AGTACTCGAA ATTTTTTGCT ATTTTGAATA AAGCACTATC
 7201 AACTTAATGA GGTTTTACTG CACATACTGT TTTGTCATTT GAAAATCTGA
 7251 AAGCACACAA AAAAAGTCAT CATTAGCCTC ACAGATCCTC ATGTGCAATA
 7301 GCTTTCCCTG AATGTTTTTA AAGGATGTAT TCTTTTGCCA AGGCCACTTC
 7351 ATATGTGCAG TAAAAAAAAA AAAAAAATCC CATTAAAGTA CAGGGGCCAG
 7401 ACAAACCAGA CTATGAAATG ATGTGATTTT AAATTTAATG ATTTTCCTTA
 7451 TTTACGAATA ATAGAGCATC TAGTTTGCCA AGAAGTCGCA GGTGCTTTGA
 7501 CTAACACACC AATTTCCACA AGCT
```

(SEQ ID NO:3)

FEATURES:
FORM 1:
Start: 3000
Exon: 3000-4523
Stop: 4524

FORM 2:
Start: 3000
Exon: 3000-3777
Intron: 3778-4107
Exon: 4108-4523
Stop: 4524

ePCR to dbSTS
no match

CHROMOSOME MAP POSITION:
Chromosome X

ALLELIC VARIANTS (SNPs):

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
4126	T	C	Exon	266(form 2)/376(form 1)	L	P

FIGURE 3, page 3 of 4

Context:

DNA
Position
4126 GAATGGAAGCCAAGGACGGCAGCCTGAAGGCCAAGGAAGGAAGCACGGGGACCAGTGAGA
GTAGTGTAGAGGCCAGGGGCAGCGAGGAGGTCAGAGAGAGCAGCACGGTGGCCAGCGACG
GCAGCATGGAGGGTAAGGAAGGCAGCACCAAAGTTGAGGAGAACAGCATGAAGGCAGACA
AGGGTCGCACAGAGGTCAACCAGTGCAGCATTGACTTGGGTGAAGATGACATGGAGTTTG
GTGAAGACGACATCAATTTCAGTGAGGATGACGTCGAGGCAGTGAACATCCCGGAGAGCC
[T,C]
CCCACCCAGTCGTCGTAACAGCAACAGCAACCCTCCTCTGCCCAGGTGCTACCAGTGCAA
AGCTGCTAAAGTGATCTTCATCATCATTTTCTCCTATGTGCTATCCCTGGGGCCCTACTG
CTTTTTAGCAGTCCTGGCCGTGTGGGTGGATGTCGAAACCCAGGTACCCCAGTGGGTGAT
CACCATAATCATCTGGCTTTTCTTCCTGCAGTGCTGCATCCACCCCTATGTCTATGGCTA
CATGCACAAGACCATTAAGAAGGAAATCCAGGACATGCTGAAGAAGTTCTTCTGCAAGGA

FIGURE 3, page 4 of 4

SEQUENCE LISTING

<110> PE CORPORATION (NY)

<120> ISOLATED HUMAN G-PROTEIN COUPLED
RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
PROTEINS, AND USES THEREOF

<130> CL000754PCT

<150> To Be Assigned
<151> 2001-11-06

<140> 09/769,741
<141> 2001-01-26

<150> 60/205,166
<151> 2000-05-18

<150> 09/638,018
<151> 2000-04-14

<160> 7

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 1527
<212> DNA
<213> Human

<400> 1
atgacgtcca cctgcaccaa cagcacgcgc gagagtaaca gcagccacac
gtgcatgccc 60
ctctccaaaa tgcccatcag cctggcccac ggcatcatcc gctcaaccgt
gctggttatc 120
ttcctcgccg cctctttcgt cggcaacata gtgctggcgc tagtgttgca
gcgcaagccg 180
cagctgctgc aggtgaccaa ccgttttatc tttaacctcc tcgtcaccga
cctgctgcag 240
atttcgctcg tggcccccctg ggtggtggcc acctctgtgc ctctcttctg
gcccctcaac 300
agccacttct gcacggccct ggttagcctc acccacctgt tcgccttcgc
cagcgtcaac 360
accattgtcg tggtgtcagt ggatcgctac ttgtccatca tccacctct
ctcctacccg 420
tccaagatga cccagcgccg cggttacctg ctcctctatg gcacctggat
tgtggccatc 480
ctgcagagca ctcctccact ctacggctgg ggccaggctg cctttgatga
gcgcaatgct 540
ctctgctcca tgatctgggg ggccagcccc agctacacta ttctcagcgt
ggtgtccttc 600
atcgtcattc cactgattgt catgattgcc tgctactccg tggtgttctg
tgcagcccgg 660
aggcagcatg ctctgctgta caatgtcaag agacacagct tggaagtgcg
agtcaaggac 720
tgtgtggaga atgaggatga agagggagca gagaagaagg aggagttcca
ggatgagagt 780
gagtttcgcc gccagcatga aggtgaggtc aaggccaagg agggcagaat
ggaagccaag 840

gacggcagcc tgaaggccaa ggaaggaagc acggggacca gtgagagtag
tgtagaggcc 900
aggggcagcg aggaggtcag agagagcagc acggtggcca gcgacggcag
catggagggt 960
aaggaaggca gcaccaaagt tgaggagaac agcatgaagg cagacaaggg
tcgcacagag 1020
gtcaaccagt gcagcattga cttgggtgaa gatgacatgg agtttggtga
agacgacatc 1080
aatttcagtg aggatgacgt cgaggcagtg aacatcccgg agagcctccc
acccagtcgt 1140
cgtaacagca acagcaaccc tcctctgccc aggtgctacc agtgcaaagc
tgctaaagtg 1200
atcttcatca tcattttctc ctatgtgcta tccctggggc cctactgctt
tttagcagtc 1260
ctggccgtgt gggtggatgt cgaaacccag gtaccccagt gggtgatcac
cataatcatc 1320
tggcttttct tcctgcagtg ctgcatccac ccctatgtct atggctacat
gcacaagacc 1380
attaagaagg aaatccagga catgctgaag aagttcttct gcaaggaaaa
gccccccgaaa 1440
gaagatagcc acccagacct gcccggaaca gagggtggga ctgaaggcaa
gattgtccct 1500
tcctacgatt ctgctacttt tccttga
1527

<210> 2
<211> 508
<212> PRT
<213> Human

<400> 2

Met	Thr	Ser	Thr	Cys	Thr	Asn	Ser	Thr	Arg	Glu	Ser	Asn	Ser	Ser	His
1				5					10					15	
Thr	Cys	Met	Pro	Leu	Ser	Lys	Met	Pro	Ile	Ser	Leu	Ala	His	Gly	Ile
			20					25					30		
Ile	Arg	Ser	Thr	Val	Leu	Val	Ile	Phe	Leu	Ala	Ala	Ser	Phe	Val	Gly
		35					40					45			
Asn	Ile	Val	Leu	Ala	Leu	Val	Leu	Gln	Arg	Lys	Pro	Gln	Leu	Leu	Gln
	50					55					60				
Val	Thr	Asn	Arg	Phe	Ile	Phe	Asn	Leu	Leu	Val	Thr	Asp	Leu	Leu	Gln
65					70					75					80
Ile	Ser	Leu	Val	Ala	Pro	Trp	Val	Val	Ala	Thr	Ser	Val	Pro	Leu	Phe
				85					90					95	
Trp	Pro	Leu	Asn	Ser	His	Phe	Cys	Thr	Ala	Leu	Val	Ser	Leu	Thr	His
			100					105					110		
Leu	Phe	Ala	Phe	Ala	Ser	Val	Asn	Thr	Ile	Val	Val	Val	Ser	Val	Asp
		115					120					125			
Arg	Tyr	Leu	Ser	Ile	Ile	His	Pro	Leu	Ser	Tyr	Pro	Ser	Lys	Met	Thr
	130					135					140				
Gln	Arg	Arg	Gly	Tyr	Leu	Leu	Leu	Tyr	Gly	Thr	Trp	Ile	Val	Ala	Ile
145					150					155					160
Leu	Gln	Ser	Thr	Pro	Pro	Leu	Tyr	Gly	Trp	Gly	Gln	Ala	Ala	Phe	Asp
				165					170					175	
Glu	Arg	Asn	Ala	Leu	Cys	Ser	Met	Ile	Trp	Gly	Ala	Ser	Pro	Ser	Tyr
			180					185					190		
Thr	Ile	Leu	Ser	Val	Val	Ser	Phe	Ile	Val	Ile	Pro	Leu	Ile	Val	Met
		195					200					205			
Ile	Ala	Cys	Tyr	Ser	Val	Val	Phe	Cys	Ala	Ala	Arg	Arg	Gln	His	Ala
	210					215					220				
Leu	Leu	Tyr	Asn	Val	Lys	Arg	His	Ser	Leu	Glu	Val	Arg	Val	Lys	Asp
225					230					235					240

```
Cys Val Glu Asn Glu Asp Glu Glu Gly Ala Glu Lys Lys Glu Glu Phe
                245                 250                 255
Gln Asp Glu Ser Glu Phe Arg Arg Gln His Glu Gly Glu Val Lys Ala
            260                 265                 270
Lys Glu Gly Arg Met Glu Ala Lys Asp Gly Ser Leu Lys Ala Lys Glu
        275                 280                 285
Gly Ser Thr Gly Thr Ser Glu Ser Ser Val Glu Ala Arg Gly Ser Glu
    290                 295                 300
Glu Val Arg Glu Ser Ser Thr Val Ala Ser Asp Gly Ser Met Glu Gly
305                 310                 315                 320
Lys Glu Gly Ser Thr Lys Val Glu Glu Asn Ser Met Lys Ala Asp Lys
                325                 330                 335
Gly Arg Thr Glu Val Asn Gln Cys Ser Ile Asp Leu Gly Glu Asp Asp
            340                 345                 350
Met Glu Phe Gly Glu Asp Asp Ile Asn Phe Ser Glu Asp Asp Val Glu
        355                 360                 365
Ala Val Asn Ile Pro Glu Ser Leu Pro Pro Ser Arg Arg Asn Ser Asn
    370                 375                 380
Ser Asn Pro Pro Leu Pro Arg Cys Tyr Gln Cys Lys Ala Ala Lys Val
385                 390                 395                 400
Ile Phe Ile Ile Ile Phe Ser Tyr Val Leu Ser Leu Gly Pro Tyr Cys
                405                 410                 415
Phe Leu Ala Val Leu Ala Val Trp Val Asp Val Glu Thr Gln Val Pro
            420                 425                 430
Gln Trp Val Ile Thr Ile Ile Ile Trp Leu Phe Phe Leu Gln Cys Cys
        435                 440                 445
Ile His Pro Tyr Val Tyr Gly Tyr Met His Lys Thr Ile Lys Lys Glu
    450                 455                 460
Ile Gln Asp Met Leu Lys Lys Phe Phe Cys Lys Glu Lys Pro Pro Lys
465                 470                 475                 480
Glu Asp Ser His Pro Asp Leu Pro Gly Thr Glu Gly Gly Thr Glu Gly
                485                 490                 495
Lys Ile Val Pro Ser Tyr Asp Ser Ala Thr Phe Pro
            500                 505
```

<210> 3
<211> 7524
<212> DNA
<213> Human

<400> 3

```
atctcacaag gctggggtaa agattcagcg gtattctctg tgtaaacagt
ttaacacagt      60
acctggcaga gagaggggga gaaagaggga gagagggaga gatgctgact
gacttacaga     120
cagacatact tcaatcaaag aatatagtgg gtaattatgt ggcttgtctt
aaatgaaatg     180
cttccaaatc ccacagggaa agcactgccg cttagctctt cctgggagct
agagaaattc     240
gtggaaaatc gttgacatta aagtcacgat tcctttctcg ggctcagggg
ctattcatac     300
tttggagggt tttcagaagt cgaggggagt cccgggccct cctatttgga
agaatcggaa     360
ggatagtgca gggtctgtgc ctgggagctg gcccgaaatc aggggctag
gaaggcacgg     420
ggtttgcacg aaaatcagcc ccggtgacaa tccccagtct gtgagctctc
ggtggactcg     480
gggccccagg ggtgggaacc cggagtgaga aaatctgagt ggggtctggt
gcacggcccc     540
ccacctgcca gccccgctcc ctgggctgcg gcggcctcgt gcgtcacggg
cgcaggcggg     600
```

ggccgagggg agggcggcac cgcaaggtga gggcaccgcg ctgccttaag
acgccgcgcg 660
cccccgctgc ctctcagagc ttgcagcgcg gcgcgctggc cggacgtgcg
ggcacagacg 720
ctgtgccccg ctcctgccgc gcccgcagcg cccgctgccc cgtgccagct
cgcagccgcc 780
cgagccccga gaggcggccg ggacgagcgc gaggagcccc gaaggacctg
cacgcggcgc 840
ccgcgacgcc acgtctgcag cagcaacagg taagcagccc ggccctactc
cgggctccgt 900
ggcccccagc acccggcagg gccccgagcc tgggagccca cgacgccccc
tctgaagcgt 960
cccgacggct gctgagctac caggcgcggt ggcccgcggt gcggcccctt
ctcgcctttc 1020
cctagccatc ctcccgccct cgcgttttct ctctcacttt ttctcgtgcg
ttcttttccc 1080
ccttcctctc cattctctcg ttttttcctc atcctttccc tcttcctctg
cttgttctct 1140
tttcttatta ttttttgctt tttctgtctg cgtctctccc agtctgtctt
gctctctcat 1200
tctctctcac tctcacgctc ccatttcttc ttttttcttt ctttctccct
ctctgttgct 1260
cacattctct ctgtctctct gtatgtatct cattctttcg ctgtctctca
tttccttttt 1320
gtctctctgc ctctcagtct ctctcgccct gcctttctga ctctctccac
caccctgcac 1380
ccctgcctct tttgcctttc tccccagctc ctttcccact cccctcggct
gtcaccgaca 1440
ccaagcagat ccggagaaag caaagaggat gccaaattgg gaatggggat
ggggctgtgg 1500
agggtgaggg taggggagct caataggcat cctctctgca ttttccagtc
tccattgcac 1560
tgctgataaa ctaagggttt tttaaccaaa gcccgcggtg gccggggcgg
caggtggttt 1620
ccacggcaaa ataaagtcca cagggagaca gagatccaaa gagacagaga
tccagccaga 1680
gaaaaaccaa ggtgaggggc catgggaaaa agatgtagag aaatagagat
gcgcacacac 1740
tcaaagatac aaagcaagtg tttagccaga gtcctggggt gacaggaaag
tgaaggacag 1800
agaaaccaac gagagagtgc acacactaaa gagatagctt gagagacagg
tgagacaagc 1860
agtagagtgt tagagatcaa tggagagaga ggcagagtaa tgcagagaga
acttaggcag 1920
ggagagacag aaaccacacg gagagacaca gagacaagag ggtccgagaa
agacagagca 1980
agcgcacacg tgccagtgag agacagaagg agacacccag agaacccgag
ggagacgcag 2040
aaacaagaga gttggagaga cacggagacg ggcagagaca cagcgaacga
cattgaacag 2100
gtgcgctagg gacccacaaa gatgaactgc tggggaaata gactgagaga
gaaatggaga 2160
gacaagagag cactagggaa agacggagac acacatgaac tagcgagcga
cacagtgaca 2220
ggcagagaca cagagagaca gaatggatca gaaagagact gagggagaga
cgggcaagac 2280
tgtgagaaag atagggagcg agagcgcaag ccctgcagag aaagtgcgct
agcgaaagag 2340
acacacacag agagactgag tgacaaacag agtggaagag ataaatgaga
gtaacaggaa 2400

agagacagcg tgcacgtgct atcgagagac ggagaaacag agacagacat
aagtgagaga 2460
caaggatgag aaagaatcag gaaacggtga gacacataca gacacacaac
ccgagagata 2520
aaggaaaaaa agtgcgctag cgagagacac tgggagaaga gacaggaaac
tgacggtcgt 2580
aaacagacag aaacagagag cgcacgcgca tgcgctagcg ctagcgacca
aaactccccg 2640
gagccagaga cagtgtgaga gacaagcaga gaaagcgcac gcacgcacgc
gccagcagga 2700
gaaacagatg agaggaaatc agagccctgg agagagacag gcagacagat
ctggagagtc 2760
cggaaaggag ccatagaagc tgcccgcact ggggatggag ccgtgcggaa
acccggggta 2820
gggggtcctg cagcgtcctt gctgggcgcg gaggcttctc cccttgacgg
gtgactaact 2880
ctgcctgcgt gtttcttttg tcaccagcat aggcactgag tgcggtctgt
gcaccccttt 2940
gccacccacc ggtgccggca ctgagcctgc aacctgtctc acgccctctg
gctgttgcca 3000
tgacgtccac ctgcaccaac agcacgcgcg agagtaacag cagccacacg
tgcatgcccc 3060
tctccaaaat gcccatcagc ctggcccacg gcatcatccg ctcaaccgtg
ctggttatct 3120
tcctcgccgc ctctttcgtc ggcaacatag tgctggcgct agtgttgcag
cgcaagccgc 3180
agctgctgca ggtgaccaac cgttttatct ttaacctcct cgtcaccgac
ctgctgcaga 3240
tttcgctcgt ggcccctgg gtggtggcca cctctgtgcc tctcttctgg
cccctcaaca 3300
gccacttctg cacggccctg gttagcctca cccacctgtt cgccttcgcc
agcgtcaaca 3360
ccattgtcgt ggtgtcagtg gatcgctact tgtccatcat ccaccctctc
tcctacccgt 3420
ccaagatgac ccagcgccgc ggttacctgc tcctctatgg cacctggatt
gtggccatcc 3480
tgcagagcac tcctccactc tacggctggg gccaggctgc ctttgatgag
cgcaatgctc 3540
tctgctccat gatctggggg gccagcccca gctacactat tctcagcgtg
gtgtccttca 3600
tcgtcattcc actgattgtc atgattgcct gctactccgt ggtgttctgt
gcagcccgga 3660
ggcagcatgc tctgctgtac aatgtcaaga gacacagctt ggaagtgcga
gtcaaggact 3720
gtgtggagaa tgaggatgaa gagggagcag agaagaagga ggagttccag
gatgagagtg 3780
agtttcgccg ccagcatgaa ggtgaggtca aggccaagga gggcagaatg
gaagccaagg 3840
acggcagcct gaaggccaag gaaggaagca cggggaccag tgagagtagt
gtagaggcca 3900
ggggcagcga ggaggtcaga gagagcagca cggtggccag cgacggcagc
atggagggta 3960
aggaaggcag caccaaagtt gaggagaaca gcatgaaggc agacaagggt
cgcacagagg 4020
tcaaccagtg cagcattgac ttgggtgaag atgacatgga gtttggtgaa
gacgacatca 4080
atttcagtga ggatgacgtc gaggcagtga acatcccgga gagcctccca
cccagtcgtc 4140
gtaacagcaa cagcaaccct cctctgccca ggtgctacca gtgcaaagct
gctaaagtga 4200

tcttcatcat catttctcc tatgtgctat ccctggggcc ctactgcttt
ttagcagtcc 4260
tggccgtgtg ggtggatgtc gaaacccagg taccccagtg ggtgatcacc
ataatcatct 4320
ggcttttctt cctgcagtgc tgcatccacc cctatgtcta tggctacatg
cacaagacca 4380
ttaagaagga aatccaggac atgctgaaga agttcttctg caaggaaaag
cccccgaaag 4440
aagatagcca cccagacctg cccggaacag agggtgggac tgaaggcaag
attgtccctt 4500
cctacgattc tgctactttt ccttgaagtt agttctaagg caaaccttga
actgtccata 4560
acacgagaaa caagaggaga tttcttttca atggacccac aattcattaa
tgccaaacca 4620
taccatttca ggcaaaggtg ttgcacacac atgctcttca ccacaaggta
gataaatata 4680
tagaagaggc aggaactggg gtctttccgt aaaagcatgg acttgaggat
tctgactgaa 4740
attttccccc caaagattat taggctctac atttcttaaa gcaacaaggg
ctatccattt 4800
tggacttgta gttggtattc tatcttttcc agagctacaa catgccaact
ttagctctga 4860
aggaaaggga agatgatgct tgtgaactta aggacttttc ggccctcggg
tcgggagctc 4920
atgggccaga gctacagctt gtgttcaact gaaagaaagg caatggacca
aatcattcat 4980
ggagcccaag aaacagaacc taatggactg atcaacatat gagccaaatt
ctgaactgaa 5040
cagccccaca gtcgggtgca aagactgtta cacaaactaa aacaaagggc
ctcctacagt 5100
tagaatctca agaaggtttc tagatcccct aaagggatcc agaaaagtag
aaggacatgt 5160
atgaaatggg aagctagtcc aagggaaaag attgagaaat aacacacatc
tggagagcta 5220
aacagttgac ttttttttcct ataaaatctt gggtttatgc atgggctgga
actgaggtca 5280
ttaagtgtaa attgtcaatt gacacaaata ttttctgtct cctgtttgaa
taatagtggg 5340
gcagaaatca tgccactatt ttacaacttc ccttatgtga ctgaattgag
atgctggtgg 5400
gaattcttca gatctctgcc aacacttctg ttttcttttg gtttgttttt
gtcaaataag 5460
ccttttttag tcaaacagta tttacagaaa aaaagaaatt caactagaag
tggcctaagt 5520
cctacaaaat tcatgatgtc actgaggaat aatttgttca tcagaaatat
attttgtgtc 5580
catgagatca tagacaataa atgtgatctc cacatgggga gcaaggaagg
cagaatgaac 5640
atttttcttc ctccaggcac acccatgtgt cttttccacc tgtggctctc
tttaaagctt 5700
ttaagctctc tgcagatgtg aaagagaaat atcagagagt cagaaatgac
aaagaggatg 5760
atttcacaat acctagaaaa catgtaacct attccaaaca gtcctaaaat
cagagcattc 5820
agatcagaca tatcctaatt aatgctgttg aaataaatca cgttgggaaa
actttaacaa 5880
tatctaaatt atccctaggg tcaattcaca ggaacattcc tcaaatccca
aaccgcaaaa 5940
taactttggg cagggatata catatacatt tctgagggca tgggccgtat
gtatgtgacc 6000

aagtaacatg gaaccaaaaa aacagtcaag ccagtgtttt tgatcctcct
acagaacaag 6060
ttaaagcaac tccagagtca accaactgtt catgcagaaa tccactgtca
atatgggtgg 6120
agggagtgtt tggttgaaaa tggttaatca ggtagttgta tgatgtaaga
tgaccatctt 6180
cagagtttag cctcattctt gtgtgattgt catgccttta ttagaactca
agtttcattt 6240
aaataaatgc ccagctcatt tatttttttt tatctcttcc tcctcacaga
tttcaacatg 6300
aacttctcaa ggggtaaaca gcaatgtatt tggactgtga ataactctgc
atgggaattt 6360
gggattgcca tgttcagaat ttaggaaagt agagggaata gaaccaaata
atatagagct 6420
gacccatcca ataaaaatac catgataacc ttattaattc caactccatt
aatttggaac 6480
ttgtagttat tcagacaagg catgggggct aaagtttacc cttacactat
catttatttt 6540
tctaccaaaa tgcataaagt gaattaacag tcataaattt gttctaccaa
tttattgcca 6600
aactcttgaa ctgacctttc cttaaaggat atctgggtga agaaatggcc
tatgtgatca 6660
atcctcctac aggggagggg cagtgccttc aggtctgttc aaattgtcaa
aggagttcaa 6720
ggtagctata gcctatcctt tgagtagaaa tgcttacttg ggtaggaaac
aggatttcaa 6780
acataaatgt ctccaaacat tgtgttaaaa ctgaacttct tgttttattt
ttaaagctca 6840
ccccccttcaa agtgtatcag agaaaattgt ttgccaaaat ctcaaatcaa
aatggaacca 6900
gaacctgtac aagtatggta agtccattta acagcacaca caacattctc
aagggcactg 6960
agattccctt tccttttga agttcctctt ttccctattt tagtcattgt
ccattatttt 7020
ggtaaattgg attcctcaaa agtgaagacc ttttgaaact atgagcctgg
aaaagaggac 7080
cttttaatta agagcattct gctttgatga cattttcttc taaatgaaca
ataaggccta 7140
tggttagctt ggagatagca agtactcgaa atttttgct attttgaata
aagcactatc 7200
aacttaatga ggttttactg cacatactgt tttgtcattt gaaaatctga
aagcacacaa 7260
aaaaagtcat cattagcctc acagatcctc atgtgcaata gctttccctg
aatgttttta 7320
aaggatgtat tcttttgcca aggccacttc atatgtgcag taaaaaaaaa
aaaaaaatcc 7380
cattaaagta caggggccag acaaaccaga ctatgaaatg atgtgatttt
aaatttaatg 7440
attttcctta tttacgaata atagagcatc tagtttgcca agaagtcgca
ggtgctttga 7500
ctaacacacc aatttccaca agct
7524

<210> 4
<211> 1197
<212> DNA
<213> Human

<400> 4
atgacgtcca cctgcaccaa cagcacgcgc gagagtaaca gcagccacac
gtgcatgccc 60

ctctccaaaa tgcccatcag cctggcccac ggcatcatcc gctcaaccgt
gctggttatc 120
ttcctcgccg cctctttcgt cggcaacata gtgctggcgc tagtgttgca
gcgcaagccg 180
cagctgctgc aggtgaccaa ccgttttatc tttaacctcc tcgtcaccga
cctgctgcag 240
atttcgctcg tggcccccctg ggtggtggcc acctctgtgc ctctcttctg
gcccctcaac 300
agccacttct gcacggccct ggttagcctc acccacctgt tcgccttcgc
cagcgtcaac 360
accattgtcg tggtgtcagt ggatcgctac ttgtccatca tccaccctct
ctcctacccg 420
tccaagatga cccagcgccg cggttacctg ctcctctatg gcacctggat
tgtggccatc 480
ctgcagagca ctcctccact ctacggctgg ggccaggctg cctttgatga
gcgcaatgct 540
ctctgctcca tgatctgggg ggccagcccc agctacacta ttctcagcgt
ggtgtccttc 600
atcgtcattc cactgattgt catgattgcc tgctactccg tggtgttctg
tgcagcccgg 660
aggcagcatg ctctgctgta caatgtcaag agacacagct tggaagtgcg
agtcaaggac 720
tgtgtggaga atgaggatga agagggagca gagaagaagg aggagttcca
ggatgagatg 780
aacatcccgg agagcctccc acccagtcgt cgtaacagca acagcaaccc
tcctctgccc 840
aggtgctacc agtgcaaagc tgctaaagtg atcttcatca tcattttctc
ctatgtgcta 900
tccctggggc cctactgctt tttagcagtc ctggccgtgt gggtggatgt
cgaaacccag 960
gtaccccagt gggtgatcac cataatcatc tggcttttct tcctgcagtg
ctgcatccac 1020
ccctatgtct atggctacat gcacaagacc attaagaagg aaatccagga
catgctgaag 1080
aagttcttct gcaaggaaaa gcccccgaaa gaagatagcc acccagacct
gcccggaaca 1140
gagggtggga ctgaaggcaa gattgtccct tcctacgatt ctgctacttt tccttga
1197

<210> 5
<211> 398
<212> PRT
<213> Human

<400> 5
Met Thr Ser Thr Cys Thr Asn Ser Thr Arg Glu Ser Asn Ser Ser His
1 5 10 15
Thr Cys Met Pro Leu Ser Lys Met Pro Ile Ser Leu Ala His Gly Ile
20 25 30
Ile Arg Ser Thr Val Leu Val Ile Phe Leu Ala Ala Ser Phe Val Gly
35 40 45
Asn Ile Val Leu Ala Leu Val Leu Gln Arg Lys Pro Gln Leu Leu Gln
50 55 60
Val Thr Asn Arg Phe Ile Phe Asn Leu Leu Val Thr Asp Leu Leu Gln
65 70 75 80
Ile Ser Leu Val Ala Pro Trp Val Val Ala Thr Ser Val Pro Leu Phe
85 90 95
Trp Pro Leu Asn Ser His Phe Cys Thr Ala Leu Val Ser Leu Thr His
100 105 110
Leu Phe Ala Phe Ala Ser Val Asn Thr Ile Val Val Val Ser Val Asp
115 120 125

```
Arg Tyr Leu Ser Ile Ile His Pro Leu Ser Tyr Pro Ser Lys Met Thr
    130                 135                 140
Gln Arg Arg Gly Tyr Leu Leu Leu Tyr Gly Thr Trp Ile Val Ala Ile
145                 150                 155                 160
Leu Gln Ser Thr Pro Pro Leu Tyr Gly Trp Gly Gln Ala Ala Phe Asp
                165                 170                 175
Glu Arg Asn Ala Leu Cys Ser Met Ile Trp Gly Ala Ser Pro Ser Tyr
            180                 185                 190
Thr Ile Leu Ser Val Val Ser Phe Ile Val Ile Pro Leu Ile Val Met
        195                 200                 205
Ile Ala Cys Tyr Ser Val Val Phe Cys Ala Ala Arg Arg Gln His Ala
    210                 215                 220
Leu Leu Tyr Asn Val Lys Arg His Ser Leu Glu Val Arg Val Lys Asp
225                 230                 235                 240
Cys Val Glu Asn Glu Asp Glu Glu Gly Ala Glu Lys Lys Glu Glu Phe
                245                 250                 255
Gln Asp Glu Met Asn Ile Pro Glu Ser Leu Pro Pro Ser Arg Arg Asn
            260                 265                 270
Ser Asn Ser Asn Pro Pro Leu Pro Arg Cys Tyr Gln Cys Lys Ala Ala
        275                 280                 285
Lys Val Ile Phe Ile Ile Ile Phe Ser Tyr Val Leu Ser Leu Gly Pro
    290                 295                 300
Tyr Cys Phe Leu Ala Val Leu Ala Val Trp Val Asp Val Glu Thr Gln
305                 310                 315                 320
Val Pro Gln Trp Val Ile Thr Ile Ile Ile Trp Leu Phe Phe Leu Gln
                325                 330                 335
Cys Cys Ile His Pro Tyr Val Tyr Gly Tyr Met His Lys Thr Ile Lys
            340                 345                 350
Lys Glu Ile Gln Asp Met Leu Lys Lys Phe Phe Cys Lys Glu Lys Pro
        355                 360                 365
Pro Lys Glu Asp Ser His Pro Asp Leu Pro Gly Thr Glu Gly Gly Thr
    370                 375                 380
Glu Gly Lys Ile Val Pro Ser Tyr Asp Ser Ala Thr Phe Pro
385                 390                 395
```

<210> 6
<211> 196
<212> PRT
<213> Human

<400> 6

```
Ser Gly Val Ile Ile Thr Gln Phe Ile Ala Ile Ile Val Ile Thr Ile
 1               5                  10                  15
Phe Val Cys Leu Gly Asn Leu Val Ile Val Val Thr Leu Tyr Lys Lys
            20                  25                  30
Ser Tyr Leu Leu Thr Leu Ser Asn Lys Phe Val Phe Ser Leu Thr Leu
        35                  40                  45
Ser Asn Phe Leu Leu Ser Val Leu Val Leu Pro Phe Val Ser Val Thr
    50                  55                  60
Ser Ser Ile Arg Arg Glu Trp Ile Phe Gly Val Val Trp Cys Asn Phe
65                  70                  75                  80
Ser Ala Leu Leu Tyr Leu Leu Ile Ser Ser Ala Ser Met Leu Thr Leu
                85                  90                  95
Gly Val Ile Ala Ile Asp Arg Tyr Tyr Ala Val Leu Tyr Pro Met Val
            100                 105                 110
Tyr Pro Met Lys Ile Thr Gly Asn Arg Ala Ser Val Met Ala Leu Val
        115                 120                 125
Tyr Ile Trp Leu His Ser Leu Ile Gly Cys Leu Pro Pro Leu Phe Gly
    130                 135                 140
Trp Ser Ser Val Glu Phe Asp Glu Phe Lys Trp Met Cys Val Ala Ala
145                 150                 155                 160
```

```
Trp His Arg Glu Pro Gly Tyr Thr Ala Phe Trp Gln Ile Trp Cys Ala
                165                 170                 175
Leu Phe Pro Phe Leu Val Met Ser Leu Val Cys Tyr Gly Phe Ile Phe
            180                 185                 190
Arg Val Ala Arg
        195
```

<210> 7
<211> 76
<212> PRT
<213> Human

<400> 7

```
Ser Gln Cys Lys Ala Leu Ile Thr Ile Leu Val Val Leu Gly Ala Phe
 1               5                  10                  15
Met Val Thr Trp Gly Pro Tyr Met Val Val Ile Ala Ser Glu Ala Leu
            20                  25                  30
Trp Gly Lys Ser Ser Val Ser Pro Ser Leu Glu Thr Trp Ala Thr Trp
        35                  40                  45
Leu Ser Phe Ala Ser Ala Val Cys His Pro Leu Ser Ile Tyr Gly Leu
    50                  55                  60
Trp Asn Lys Thr Val Arg Lys Glu Leu Leu Gly Met
65                  70                  75
```

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 August 2002 (01.08.2002)

PCT

(10) International Publication Number
WO 02/059151 A3

(51) International Patent Classification[7]: C07K 14/705, 16/28, C12N 15/12, 5/10, C12Q 1/68, A01K 67/027, A61K 38/17, G01N 33/566, 33/68

(21) International Application Number: PCT/US01/42978

(22) International Filing Date:
6 November 2001 (06.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/769,741 26 January 2001 (26.01.2001) US

(71) Applicant: PE CORPORATION (NY) [US/US]; 761 Main Avenue, Norwalk, CT 06859 (US).

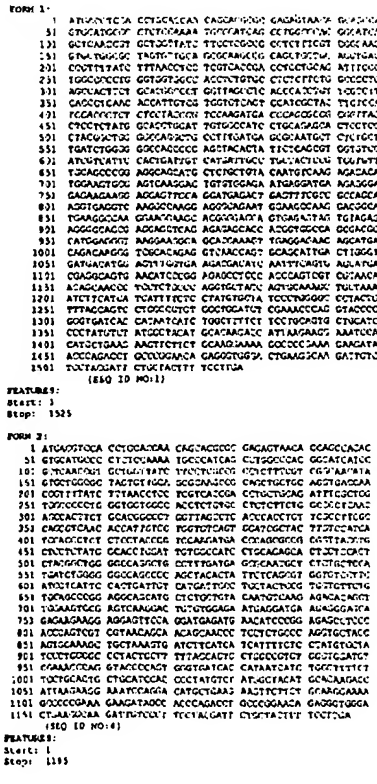
(72) Inventors: LI, Zhenya; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). CRAVCHIK, Anibal; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). MILSHINA, Natalia; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). WEI, Ming-Hui; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). KETCHUM, Karen, A.; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). DI FRANCESCO, Valentina; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). BEASLEY, Ellen, M.; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US).

(74) Agent: CELERA GENOMICS; Millman, Robert, A., Chief Intellectual Property Counsel, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,

[Continued on next page]

(54) Title: ISOLATED HUMAN G-PROTEIN COUPLED RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR PROTEINS, AND USES THEREOF



(57) Abstract: The present invention provides amino acid sequences of peptides that are encoded by genes within the human genome, the GPCR peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthlogs and paralogs of the GPCR peptides and methods of identifying modulators of the GPCR peptides.

WO 02/059151 A3

LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) **Date of publication of the international search report:**
17 April 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/42978

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/705 C07K16/28 C12N15/12 C12N5/10 C12Q1/68
A01K67/027 A61K38/17 G01N33/566 G01N33/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, EMBL, WPI Data, PAJ, MEDLINE, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 98 20040 A (INCYTE PHARMA INC ;AU YOUNG JANICE (US); GOLI SURYA K (US); GUEGLE) 14 May 1998 (1998-05-14) SEQ ID NO 1 figure 1 ---	1-16,19
A	HIRASAWA AKIRA ET AL: "Cloning, functional expression and tissue distribution of human alpha-1C-adrenoceptor splice variants." FEBS LETTERS, vol. 363, no. 3, 1995, pages 256-260, XP002216676 ISSN: 0014-5793 figure 1 --- -/--	1-16,19

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 October 2002	31/10/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Lanzrein, M

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 01 62797 A (PARODI LUIS A ;LIND PETER (SE); UPJOHN CO (US); VOGELI GABRIEL (US) 30 August 2001 (2001-08-30) SEQ ID NOs 52, 112 page 89 -page 90	1-16,19
P,X	WO 01 81410 A (LIND PETER ;SEJLITZ TORSTEN (SE); UPJOHN CO (US); VOGELI GABRIEL () 1 November 2001 (2001-11-01) SEQ ID NOs 1, 2	1-16,19
P,X	WO 01 36471 A (ARENA PHARMACEUTICALS INC ;CHEN RUOPING (US); DANG HUONG T (US); L) 25 May 2001 (2001-05-25) SEQ ID NOs 15, 16	1-16,19
E	EP 1 243 648 A (MITSUBISHI PHARMA CORP) 25 September 2002 (2002-09-25) SEQ ID NOs 21, 26	1-16,19
E	WO 01 88126 A (BAYER AG ;RAMAKRISHNAN SHYAM (US)) 22 November 2001 (2001-11-22) SEQ ID NOs 2, 3 figures 2,3	1-16,19

INTERNATIONAL SEARCH REPORT

national application No.
PCT/US 01/42978

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 17, 18
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

International Application No. PCT/US 01 42978

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 17, 18

Present claims 17, 18 relate to an extremely large number of possible compounds/methods. Support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT is not found for any such compounds/methods claimed. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the claimed scope is impossible.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/42978

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9820040 A	14-05-1998	US 5817480 A	06-10-1998
		AU 5169598 A	29-05-1998
		EP 0941249 A1	15-09-1999
		JP 2001503628 T	21-03-2001
		WO 9820040 A1	14-05-1998
WO 0162797 A	30-08-2001	AU 4165801 A	03-09-2001
		AU 5787501 A	09-07-2001
		EP 1242454 A2	25-09-2002
		WO 0148015 A2	05-07-2001
		WO 0162797 A2	30-08-2001
		US 2002062013 A1	23-05-2002
		AU 4165501 A	03-09-2001
WO 0181410 A	01-11-2001	AU 5723101 A	07-11-2001
		WO 0181410 A2	01-11-2001
WO 0136471 A	25-05-2001	AU 1769601 A	30-05-2001
		EP 1242448 A2	25-09-2002
		WO 0136471 A2	25-05-2001
EP 1243648 A	25-09-2002	AU 2230301 A	09-07-2001
		EP 1243648 A1	25-09-2002
		AU 2230401 A	09-07-2001
		WO 0148188 A1	05-07-2001
		WO 0148189 A1	05-07-2001
		AU 4467301 A	08-10-2001
		WO 0173023 A1	04-10-2001
WO 0188126 A	22-11-2001	AU 7051901 A	26-11-2001
		WO 0188126 A2	22-11-2001